THE UNIVERSITY OF ALBERTA

ROBUST ESTIMATES OF LOCATION: INTERMEDIARIES BETWEEN SAMPLE MEAN AND MEDIAN

by

© LADISLAV Z. FERENCZI

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

EDMONTON, ALBERTA

FALL, 1972

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "ROBUST ESTIMATES OF LOCATION: INTERMEDIARIES BETWEEN SAMPLE MEAN AND MEDIAN" submitted by LADISLAV Z. FERENCZI in partial fulfilment of the requirements for the degree of Master of Science.

ABSTRACT

In Chapter I the problem of robust estimation of location parameter is introduced and an outline of results of J.W. Tukey - who first investigated this problem - is given.

In Chapter II an important method of robust estimation of location based on rank tests is shown, due to J.L. Hodges Jr. and E.L. Lehmann.

In Chapter III we present the minimax approach toward the theory of robust estimation of location parameter, due to P.J. Huber. Also an important class of estimates based on minimal principle is introduced and a unique most robust estimate for the location parameter - according to an accepted measure of robustness - is defined.

In Chapter IV some new estimates of location parameter are presented which in a sense lie between sample mean and sample median, and which are based on minimal principle. According to an accepted measure of robustness, two most robust estimates for location parameter in two classes of estimates are defined, and some of their properties are investigated.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Dr. R. Ponnapalli for suggesting to me the topic of robust estimation and for his patient and enthusiastic guidance during the research and preparation of this thesis. I should also like to thank Miss June Talpash for typing the thesis.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

ROBUST ESTIMATION OF LOCATION PARAMETER

I.1 Introduction

The problem of the theory of point estimation is to develop methods for finding estimates on the basis of sample values $x_1,\ldots,x_n$ i.e., on the basis of observed values of the random variables $X_1,\ldots,X_n$ whose distribution function is assumed to be $F_\theta(x)$, which, according to certain criteria, are best for estimating the unknown parameter θ of the assumed underlying distribution function $F_\theta(x)$.

The assumption of normality of the distribution function $F_\theta(x)$ is frequent in classical statistical methods.

In reality we never have a complete knowledge about the true underlying distribution function, and hence the assumed and the true underlying distribution functions may differ. Therefore it is necessary to know how an estimate, obtained by a certain method will perform, if the assumed underlying distribution is slightly changed or replaced by a different one.

It seems to be a desirable property of an estimate to have stable performance, at least when we allow for a small change in the assumed underlying distribution function. This was recognized in the past, but until recent times not enough attention was paid to this basic problem. The theory of robust estimation deals with this problem and tries to exhibit estimates, whose performance is relatively stable not only under one set of circumstances, as it is for example in the case when we assume only a fixed underlying distribution function

$F_\theta(x)$ in the classical problem of point estimation of a single unknown parameter θ.

The problem of robust estimation of a location is to develop methods for finding estimates, on the basis of observed values $x_1,\ldots,x_n$ of the random variables $X_1,\ldots,X_n$ having distribution function $F_\theta(x) = F(x-\theta)$, which according to certain criteria are stable (robust) estimates of the unknown location parameter θ of the distribution function $F_\theta(x) = F(x-\theta)$, when the prototype distribution function $F(x)$ is assumed to be known only approximately.

In case of a symmetric prototype distribution function i.e., if $F(x)+F(-x) = 1$, the center of symmetry of the distribution $F(x-\theta)$ is considered as the unknown location parameter θ - a natural quantity to estimate in this situation.

Thus the problems of the robust point estimation are:

(a) to choose an appropriate model of indeterminacy which will play the role of the prototype distribution function of the classical theory of point estimation;

(b) to define a most robust estimate in a reasonable class of estimates.

I.2 Outline of Tukey's Results

The shortcomings of the classical methods of estimation - the assumption of normality and consequently their vulnerability to gross errors were investigated by J.W. Tukey and the Statistical Research Group in Princeton in the late forties.

A survey paper by Tukey [14] published in 1960 is considered as a starting point in the theory of robust estimation. Tukey con-

sidered a model of indeterminacy with the prototype distribution function - the so-called contaminated normal distribution of the form:

$$F_{\varepsilon,h}(x) = (1-\varepsilon) \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt + \varepsilon \int_{-\infty}^{x} \frac{1}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t}{h})^2} dt ,$$

where $\varepsilon \in [0,1]$ is the fraction of contamination and h is the scale ratio of the two component normal distribution functions. In his investigations a scale ratio $h = 3$ has been used throughout.

The purpose of the contaminated normal distribution function was to replace the exact normal distribution function in the classical problem of estimation of the location and the scale parameters.

Tukey has shown, that the classical estimate of location parameter μ of the distribution function $F_{\varepsilon,3}(x-\mu)$ - the arithmetic mean, and the classical estimate of scale parameter σ of the distribution $F_{\varepsilon,3}(x/\sigma)$ - the standard deviation, both have unsatisfactory aspects, namely their variances exploded even for a very small amount of contamination fraction ε.

Tukey proposed estimates for location and scale, which are more robust than the classical estimates. The asymptotic efficiency and the asymptotic effective variance have been used as a criteria of robustness of an estimate in location and scale problems respectively.

As an estimate of location parameter μ, the α-trimmed mean (the arithmetic mean of those observations, which remain, when the α% lowest and α% highest have been set aside), $\bar{X}_{\alpha}$,

$$\bar{X}_\alpha = \{n-2[\alpha n]\}^{-1} \sum_{i=[\alpha n]+1}^{n-[\alpha n]} Y_{(i)} \quad ,$$

was proposed, where $0 \leq \alpha < \frac{1}{2}$, and $Y_{(1)},\ldots,Y_{(n)}$ denote the order statistics of the sample $X_1,\ldots,X_n$ from the distribution $F_{\varepsilon,3}(x-\mu)$.

As an estimate of scale parameter σ, the mean deviation was proposed for smaller samples and in larger samples, the α-truncated standard deviation (the standard deviation of those observations, which remain, when the α% lowest and α% highest have been set aside), s_α,

$$s_\alpha = \{\{n-2[\alpha n]-1\}^{-1} \sum_{i=[\alpha n]+1}^{n-[\alpha n]} (Y_{(i)}-\bar{X}_\alpha)^2\}^{\frac{1}{2}} \quad ,$$

was proposed, where $0 \leq \alpha < \frac{1}{2}$ and $Y_{(1)},\ldots,Y_{(n)}$ denote the ordered statistics of the sample $X_1,\ldots,X_n$ from the distribution $F_{\varepsilon,3}(x/\sigma)$.

Tukey showed, that the α-trimmed mean $\bar{X}_\alpha$ is asymptotically normal, if the underlying prototype distribution function $F(x)$ is symmetric and has a density, which is continuous and strictly positive on $\{x : 0 < F(x) < 1\}$. Thus

$$L[n^{\frac{1}{2}}(\bar{X}_\alpha-\mu)] \to N(0,\sigma^2(\alpha)), \quad \text{as} \quad n \to \infty \quad ,$$

where

$$\sigma^2(\alpha) = (1-2\alpha)^{-2}[\int_{x(\alpha)}^{x(1-\alpha)} t^2 dF(t) + 2\alpha x^2(\alpha)] \quad ,$$

and $x(\alpha)$ is the α-quantile of $F(x)$.

CHAPTER II

EFFICIENCY-ROBUST ESTIMATES OF LOCATION

II.1 Outline of Results of Hodges and Lehmann

In this chapter we shall present an important method of estimation due to Hodges and Lehmann [5] which leads to robust estimates of location. Their approach is one of the first attempts to deal with a weak point of some classical methods of the theory of estimation - the assumption of normality and consequently their vulnerability to gross errors. They realized that for the problems of point estimation the methods successful in the corresponding testing problems could be applied. The rank tests such as the two Wilcoxon tests or the Kruskal-Wallis H-test have more robust powers against gross errors than the t- and F-tests, and their efficiency loss is quite small even in the rare case in which the suspicion of possibility of gross errors is unfounded. The method of estimation of location or shift parameters proposed by Hodges and Lehmann is based on rank test statistics such as the Wilcoxon or normal scores statistics, which are successful in providing robust power for the corresponding testing problems.

In the following sections we shall summarize the results of Hodges and Lehmann.

II.2 Point Estimates Based on Test Statistics

Let $X_1,\ldots,X_m$; $Y_1,\ldots,Y_n$ be independent random variables with distributions

$$(2.1) \qquad P\{X_i \leq u\} = F(u); \quad P\{Y_j \leq u\} = F(u-\Delta) \quad .$$

The random variables $X_1,\ldots,X_m$; $Y_1-\Delta,\ldots,Y_n-\Delta$, are then i.i.d. random variables, since the random variables $Y_1-\Delta,\ldots,Y_n-\Delta$, are obtained by shifting the Y-sample Δ to the left.

The idea is to estimate Δ by the amount of shift needed to align as closely as possible the two sets $(X_1,\ldots,X_m)$, $(Y_1-\Delta,\ldots,Y_n-\Delta)$. The definition of alignment could be given with reference to the Wilcoxon statistic, by defining the two sets to be aligned, if half of non-zero differences $(Y_j-\Delta)-X_i$ are positive and half negative. There is either a unique such value Δ, which could then serve as an estimate, or an interval of such values; in the later case, the midpoint of this interval provides a natural estimate.

More generally, if a test of hypothesis $\Delta = 0$ is based on a statistic, whose distribution is symmetric about a point μ, the two sets could be defined in alignment, when giving to the test statistic, the value μ. Let us assume now, that $F \in \mathcal{F}_0$ or $F \in \mathcal{F}_1$, where

$$\mathcal{F}_0 = \{\text{set of all continuous distributions}\} ,$$

$$\mathcal{F}_1 = \{\text{set of all continuous distributions, symmetric about zero}\} .$$

Consider a test statistic

$$h(X_1,\ldots,X_m,Y_1,\ldots,Y_n)$$

for the hypothesis $H : \Delta = 0$ against the alternative $\Delta \neq 0$. We

shall assume throughout, that h satisfies:

(A) $h(x_1,\ldots,x_m,y_1+a,\ldots,y_n+a)$ is a nondecreasing function of a for all x and y,

(B) when $\Delta = 0$, the distribution of $h(X_1,\ldots,X_m;Y_1,\ldots,Y_n)$ is symmetric about a fixed point μ (independent of F), (i) for all $F \in \mathcal{F}_0$, or (ii) for all $F \in \mathcal{F}_1$.

We shall use the following abbreviations in the notation: $x = (x_1,\ldots,x_m)$; $y = (y_1,\ldots,y_n)$; $x < x'$ means that the inequality holds for each coordinate; if a is a real number, then $x+a = (x_1+a,\ldots,x_m+a)$. The notation $P_o\{.\}$ will mean, that the probability in question is being computed for the case $\Delta = 0$.

<u>Definition 1:</u> $\hat{\Delta} = \frac{1}{2}(\Delta^* + \Delta^{**})$, where

$$\Delta^* = \sup \{\Delta : h(x,y-\Delta) > \mu\} \quad ,$$

and

$$\Delta^{**} \inf \{\Delta : h(x,y-\Delta) < \mu\} \quad .$$

Then $\hat{\Delta}$ is proposed for the two sample problem as an estimate of shift parameter Δ, for a suitable function h.

In the case of one sample problem, suppose $Z_1,\ldots,Z_N$ are independent random variables with common distribution

$$P\{Z_i \leq u\} = F(u-\theta) \quad ,$$

where F is continuous and symmetric about zero, i.e., $F \in \mathcal{F}_1$.

Similarly as in the two sample problem we will base an estimate on a test statistic $h(Z_1,\ldots,Z_N)$ for the hypothesis $\theta = 0$

against the alternative $\theta \neq 0$. We shall again assume throughout, that h satisfies:

(C) $h(z_1+a,\ldots,z_N+a)$ is a nondecreasing function of a for each z.

(D) for $\theta = 0$, the distribution of h is symmetric about a fixed point μ (independent of F) for all $F \in \mathcal{F}_1$.

If μ is the median of $h(Z)$, when $\theta = 0$, we define

Definition 2: $\hat{\theta} = \frac{1}{2}(\theta^* + \theta^{**})$, where

$$\theta^* = \sup\{\theta : h(z-\theta) > \mu\} ,$$

and

$$\theta^{**} = \inf\{\theta : h(z-\theta) < \mu\} .$$

Then $\hat{\theta}$ is proposed for the one sample problem as an estimate of the location parameter θ.

II.3 Estimates Based on Rank Tests

An important class of rank statistics for the two-sample problem is given by

$$h(x,y) = \sum_{j=1}^{n} E_\Psi[V^{(s_j)}] , \tag{3.1}$$

where $s_1,\ldots,s_n$ denote the ranks of $y_1,\ldots,y_n$ in the combined sample and where $V^{(1)} < \ldots < V^{(m+n)}$ denote an ordered sample of size $m+n$ from a distribution Ψ.

The function h defined by (3.1) satisfies requirement (A).

Conditions under which h satisfies requirement (B) are given in the following lemma, in which h is not assumed to satisfy (3.1).

<u>LEMMA 1</u>: The distribution of $h(X,Y)$ is symmetric about μ, if any of the following three conditions hold.

(i) h is a function only of the ranks and satisfies

$$h(x,y) + h(-x,-y) = 2\mu \qquad (\text{a.e.}P_o) \tag{3.2}$$

(ii) the sample sizes m and n are equal and h satisfies

$$h(x,y) + h(y,x) = 2\mu \qquad (\text{a.e.}P_o) \tag{3.3}$$

(iii) the distribution F is symmetric about zero and h satisfies (3.2).

Conditions under which the function h defined by (3.1) satisfies (3.2) or (3.3) are given in the following lemma.

<u>LEMMA 2</u>: Let h be defined by (3.1). Then

(i) if Ψ is symmetric about b, the function h satisfies (3.2) with $\mu = n \cdot b$

(ii) if $m = n$ and b denotes the expectation of Ψ the function h satisfies (3.3) with $\mu = \frac{1}{2}(m+n)b$.

It follows from Lemmas 1 and 2, that a function h given by (3.1) satisfies condition (B)(i) of the preceeding section if either Ψ is symmetric or the two-sample sizes m and n are equal.

Among the statistics given by (3.1) and satisfying (B)(i) we shall be interested in the Wilcoxon statistic and the normal score statistic, obtained by taking for Ψ a rectangular or normal dis-

tribution respectively.

Suppose now the sample sizes are equal, i.e., $m = n$. Denote the sample mean by $\bar{x} = \frac{1}{m}(x_1+\ldots+x_m)$, and the sample median by

$$\tilde{x} = \text{med } x = \begin{cases} x^{(k+1)} & \text{if} \quad m = 2k+1 \\ \frac{1}{2}(x^{(k)} + x^{(k+1)}) & \text{if} \quad m = 2k \ , \end{cases}$$

where $x^{(1)} < \ldots < x^{(m)}$ denote the ordered x's. We can see, that $h(x,y) = \bar{y}-\bar{x}$ and $h(x,y) = \tilde{y}-\tilde{x}$ both satisfy (3.3) with $\mu = 0$ and $\Delta^{*} = \Delta^{**} = h$. The estimates for shift are therefore $\bar{Y}-\bar{X}$ and $\tilde{Y}-\tilde{X}$, respectively.

If in addition to (3.3) we impose on h the condition

$$(3.4) \qquad h(x,y+a) = h(x,y)+a \qquad \text{for all real} \quad a \ ,$$

then we can assume without loss of generality that $\mu = 0$ since the function $h'(x,y) = h(x,y-\mu)$ satisfies (3.3) with $\mu = 0$. Condition (3.4) then implies $\Delta^{*} = \Delta^{**} = h$, since for example

$$\Delta^{**}(x,y) = \inf\{\Delta : h(x,y-\Delta) < \Delta\} = \inf\{\Delta : h(x,y) < \Delta\} = $$
$$= h(x,y) \quad .$$

Suppose now, that $h(x,y)$ is the test statistic of the Wilcoxon two-sample test in the Mann-Whitney form, i.e., $h(x,y)$ is the number of pairs (i,j) such that $x_i < x_j$ $(1 \leq i \leq m;\ 1 \leq j \leq n)$.

This is equivalent to the test based on (3.1) with Ψ, the rectangular distribution on $(0,1)$. The values of function h, which satisfies requirement (B)(i) are the integers $0,1,\ldots,m.n$. To find an explicit expression for estimate $\hat{\Delta}$ of Definition 1,

based on $h(x,y)$, denote by $W^{(1)},\dots,W^{(m.n)}$ the ordered differences Y_j-X_i. Suppose first $m.n$ is odd, $mn = 2k+1$, say. Then $\mu = k + \frac{1}{2}$, and

$$\begin{aligned}\Delta^{**} &= \inf\{\Delta : h(x,y-\Delta) < k+\tfrac{1}{2}\} = \\ &= \inf\{\Delta : \text{fewer than } k+\tfrac{1}{2} \text{ of the differences } Y_j-X_i \text{ exceed } \Delta\} = \\ &= \inf\{\Delta : W^{(k+1)} \leqq \Delta\} = W^{(k+1)} .\end{aligned}$$

Similarly

$$\begin{aligned}\Delta^{*} &= \sup\{\Delta : \text{more than } k+\tfrac{1}{2} \text{ of the differences } Y_j-X_i \text{ exceed } \Delta\} = \\ &= \sup\{\Delta : W^{(k+1)} \geqq \Delta\} = W^{(k+1)} .\end{aligned}$$

Hence $\hat{\Delta} = W^{(k+1)}$. Suppose now $m.n$ is even, $m.n = 2k$, say. Then

$$\begin{aligned}\Delta^{**} &= \inf\{\Delta : W^{(k+1)} \leqq \Delta\} = W^{(k+1)} \\ \Delta^{*} &= \sup\{\Delta : W^{(k)} > \Delta\} = W^{(k)} ,\end{aligned}$$

hence

$$\hat{\Delta} = \frac{1}{2}\,[W^{(k)} + W^{(k+1)}] .$$

Thus

$$\hat{\Delta} = \text{med}\,[Y_j-X_i] , \tag{3.5}$$

is the median of $m.n$ differences of Y_j-X_i. Similar is the estimate in the case of normal scores. Next we shall consider the case of one-sample problem.

Let $Z_1,\ldots,Z_N$ be independent identically distributed random variables with distribution $F(u-\theta)$, where F is symmetric about zero. Let $s_1,\ldots,s_n$ denote the ranks of the positive Z's among N absolute values $|Z_1|,\ldots,|Z_N|$. Here n is a random variable, which for $\theta = 0$ has the binomial distribution $B(N; \frac{1}{2})$.

A class of rank tests is based on the test statistic

$$h(z) = \sum_{j=1}^{n} E_\Psi[V^{(s_j)}] \quad , \tag{3.6}$$

where $V^{(1)} < \ldots < V^{(N)}$ denote the ordered absolute values of a sample of size N from a distribution Ψ. The function h, given by (3.6) satisfies requirement (C) of the previous section. The function h satisfies requirement (D) if

$$h(z) + h(-z) = 2\mu \qquad \text{(a.e. } P_o\text{)} \tag{3.7}$$

The following lemma states, that the requirement (D) is in fact satisfied for any function h defined by (3.6)

<u>LEMMA 3</u>: If h is given by (3.6) and $\theta = 0$, the distribution of h is symmetric about $\mu = \frac{1}{2} N.E_\psi|Z_1|$ for all $F \in \mathcal{F}_1$.
An important case is again the Wilcoxon statistic, which corresponds to the choice of rectangular distribution for Ψ in (3.6).

To find an explicit expression for estimate $\hat{\theta}$ of Definition 2, based on $h(z)$ we shall use an equivalent form of test statistic $h(z)$ due to Tukey [13], **namely:**

(3.8) $\quad h(z) =$ Numbers of pairs (i,j) with $1 \leq i \leq j \leq N$ such that $z_i + z_j > 0$.

The possible values of h are integers $0,1,\ldots,\frac{1}{2}N(N+1)$. Let $W^{(1)},\ldots,W^{(K)}$ be the $K = \frac{1}{2}N(N+1)$ averages $\frac{1}{2}(z_i+z_j)$ with $i \leq j$. Then similarly as in the case of two-sample problem

$$(3.9) \qquad \hat{\theta} = \text{med}\left[\frac{z_i+z_j}{2}\right] ,$$

the median of $\frac{N(N+1)}{2}$ averages $\frac{z_i+z_j}{2}$.

Other estimates are obtained by taking for h a function, that satisfies (3.7) and the following translation invariance requirement:

$$(3.10) \qquad h(z+a) = h(z)+a \qquad \text{for all real } a .$$

As in the corresponding case of the two-sample problem we can assume without loss of generality that $\mu = 0$ and then $\hat{\theta}(z) = h(z)$. Examples of this are functions (i) $h(z) = \bar{z}$ and (ii) $h(z) = \tilde{z}$. For the proofs of the above lemmas see [5], pp. 601-603.

II.4 Properties of Estimates

A) Small Sample Properties of Estimates

The estimates $\hat{\Delta}$ and $\hat{\theta}$ of a shift or location parameter are translation invariant and approximately median unbiased.

The following theorems give conditions under which the distribution of $\hat{\Delta}$ and $\hat{\theta}$ are symmetric so that in particular the estimates are unbiased.

THEOREM 1: The distribution of the estimate $\hat{\Delta}$ of Definition 1 is symmetric about Δ if either one of the following conditions hold:

(i) the distribution F defined in (2.1) is symmetric and h satisfies (3.2) and the invariance relation

$$h(x+a,y+a) = h(x,y) \quad \text{for all} \quad a \quad . \tag{4.1}$$

(ii) the two-sample sizes m and n are equal, and h satisfies (3.3) and (4.1).

<u>Corollary:</u> If h is given by (3.1), then the distribution of $\hat{\Delta}$ is symmetric about Δ if either one of the following conditions holds:

(i) the distribution F and Ψ are symmetric

(ii) the sample sizes m and n are equal.

<u>THEOREM 2:</u> The distribution of the estimate $\hat{\theta}$ of Definition 2 is symmetric about θ if

(i) F is symmetric about zero and h satisfies (3.7) and hence in particular if

(ii) h is given by (3.6).

For the proofs of the above theorems see [5], pp. 605-607.

B) Asymptotic Properties of Estimates

The following theorems will be concerned with asymptotic properties of estimates $\hat{\Delta}$ and $\hat{\theta}$. For that purpose we shall use the following notation.

For the two-sample problem, let $m(N)$, $n(N)$ for $N = 1,2,\ldots,$ be a sequence of pairs of sample sizes tending to infinity in such a way, that $m(N)/N \to \lambda$, say, and let Δ_N be a sequence of values of the parameter Δ. Also, for the one-sample problem consider the sequence of sample sizes $N = 1,2,\ldots,$ and let θ_N be a sequence of

values of θ. In both cases we shall indicate the dependence of h and μ on N by writing h_N and μ_N.

THEOREM 3: Let $a, c_1, c_2, \ldots,$ be real constants, and let

$$\Delta_N = -\frac{a}{c_N} \quad \text{or} \quad \theta_N = -\frac{a}{c_N} .$$

Let G be the continuous distribution function of a random variable with mean zero and unit variance, and suppose

$$\lim_{N \to \infty} P_N\{c_N(h_N-\mu_N) \leq u\} = G(\frac{u+aB}{A}) ,$$

where P_N indicates, that the probability is computed for the parameter values Δ_N or θ_N and where h_N stands for

$$h_N(X_1,\ldots,X_{m(N)};Y_1,\ldots,Y_{n(N)}) \quad \text{or} \quad h_N(Z_1,\ldots,Z_N) .$$

Then for any fixed Δ and θ,

$$(4.2) \qquad \lim_{N \to \infty} P_\Delta\{c_N(\hat{\Delta}_N-\Delta) \leq a\} = G(\frac{aB}{A}) ,$$

or

$$(4.3) \qquad \lim_{N \to \infty} P_\theta\{c_N(\hat{\theta}_N-\theta) \leq a\} = G(\frac{aB}{A}) .$$

Consider now test statistics given by (3.1). Then from the results of Chernoff and Savage [2], (1958) or Puri [12], (1963) Theorem 7.1 under suitable regularity conditions on Ψ, $N^{\frac{1}{2}}[h_N(X,Y)-\mu_N]$ satisfies the assumptions of Theorem 3, with G, the standard normal distribution, and with A and B given by

$$(4.4) \qquad A^2 = \lambda(1-\lambda)[\int_0^1 J^2(u)du - (\int_0^1 J^2(u)du)^2] ,$$

and

$$B = \lambda(1-\lambda) \int \{\frac{d}{dx} (J[F(x)])\} dF(x) , \tag{4.5}$$

where $J = \Psi^{-1}$. This together with Theorem 3 gives:

THEOREM 4: If h is given by (3.1) with Ψ satisfying the assumptions of Theorem 7.1 of Puri, (1963), and if $m(N)/N \to \lambda$ as $N \to \infty$, then $N^{\frac{1}{2}}(\hat{\Delta}_N - \Delta)$ has a limiting normal distribution with mean zero and variance A^2/B^2, where A and B are given by (4.4) and (4.5).

For particular case Ψ being the rectangular distribution on $(0,1)$ we have $A^2 = \lambda(1-\lambda)$, $B = \lambda(1-\lambda) \int f^2(x)dx$, where $f(x)$ if the density of $F(x)$. The asymptotic variance of $N^{\frac{1}{2}}(\hat{\Delta}-\Delta)$ in this case is given by:

$$1/[12\lambda(1-\lambda)(\int f^2(x)dx)^2] . \tag{4.6}$$

The following theorem shows, that under suitable regularity conditions, the estimates $\hat{\Delta}$ and $\hat{\theta}$ have desirable efficiency properties.

THEOREM 5: Let $\hat{\Delta}_N$ and $\hat{\Delta}_N'$ (or $\hat{\theta}_N$ and $\hat{\theta}_N'$) be estimates of Δ (or θ) based on sequences of test statistics h_N and h_N' satisfying the assumptions of Theorem 3 for the same limiting distribution G. Then the asymptotic relative efficiency of $\hat{\Delta}_N'$ relative to $\hat{\Delta}_N$ (or of $\hat{\theta}_N'$ relative to $\hat{\theta}_N$) in the sense of reciprocal ratio of asymptotic variances, is the same as the corresponding Pitman efficiency of the two sequences of tests based on h_N' and h_N provided the latter exists and $c_N = c_N' = N^{\frac{1}{2}}$. For the proofs of

the above theorems see [5], pp. 608-610.

It follows from this theorem, that the efficiency of the estimates med $(Y_j - X_i)$ and med $((Z_i + Z_j)/2)$ relative to the classical estimates $\bar{Y} - \bar{X}$ and $\bar{Z}$ respectively is $12\sigma^2 (\int f^2(x)dx)^2$, which in the case of normal F is $\frac{3}{\pi} \doteq 0.955$.

It is interesting to compare this value with the corresponding values for small N. For $N = 1$ and 2, $N = \bar{Z}$ so that the efficiency is 1. For $N = 3$

$$\hat{\theta} = \text{med}\{Z_1, Z_2, Z_3, \frac{Z_1+Z_2}{2}, \frac{Z_1+Z_3}{2}, \frac{Z_2+Z_3}{2}\} .$$

Let the ordered Z's be denoted by $Z^{(1)} < Z^{(2)} < Z^{(3)}$. Then

$$Z^{(1)} < \frac{Z^{(1)}+Z^{(2)}}{2} < Z^{(2)} < \frac{Z^{(2)}+Z^{(3)}}{2} < Z^{(3)}$$

and

$$Z^{(1)} < \frac{Z^{(1)}+Z^{(2)}}{2} < \frac{Z^{(1)}+Z^{(3)}}{2} < \frac{Z^{(2)}+Z^{(3)}}{2} < Z^{(3)} .$$

From these inequalities it follows, that $\hat{\theta}_2$ = average of $Z^{(2)}$ and $\frac{Z^{(1)}+Z^{(3)}}{2}$, so that

$$\hat{\theta}_3 = \frac{1}{4}[Z^{(1)} + 2Z^{(2)} + Z^{(3)}] .$$

From a table of the covariances of normal order statistics, the efficiency of $\hat{\theta}_3 = 0.979$.

Generally, for any F with bounded density f, the value $12\sigma^2(\int f^2(x)ds)^2 \geq 0.864$, hence the estimate based on Wilcoxon test is efficiency robust.

CHAPTER III

MINIMAX APPROACH FOR ROBUSTNESS

III.1 Outline of Huber's Results

In this chapter we shall present the approach of P.J. Huber [6] toward a theory of robust estimation. For the problem of estimating a single location parameter, he developed a general method for obtaining estimates based on minimal principle.

In his work, he considered two models of indeterminacy. In the first model the prototype distribution $F_\varepsilon(x)$ was given by

$$F_\varepsilon(x) = (1-\varepsilon)\Phi(x) + \varepsilon H(x) \quad ,$$

where $0 \leq \varepsilon < 1$ is a known number, $\Phi(x)$ is the standard normal distribution, and H is an unknown contaminating distribution. (This model of indeterminacy is a generalization of the model studied by Tukey.) In the second model $F_\varepsilon(x)$ was given by

$$\sup_{-\infty<x<\infty} |F_\varepsilon(x)-\Phi(x)| \leqq \varepsilon \quad .$$

Under suitable regularity conditions, asymptotic normality of the estimates based on the minimal principle was shown and an explicit expression for the asymptotic variance was given.

Using a minimax approach Huber showed that for the first model of indeterminacy there exists a unique most robust estimate T_n based on minimal principle, if as a measure of robustness of an estimate we accept the inverse of the supermum of the asymptotic variance over the set of all contaminated distribution functions.

The most robust estimate T_n of location is defined by

$$\sum_{i=1}^{n} \rho(X_i - T_n) = \text{minimum} \quad ,$$

where

$$\rho(t) = \frac{1}{2} t^2 \qquad \text{for} \quad |t| < k$$
$$= k|t| - \frac{1}{2} k^2 \qquad \text{for} \quad |t| \geq k$$

with k depending on ε.

It was shown in [8], that the above estimate is asymptotically equivalent to the trimmed mean $\bar{X}_\alpha$ introduced by Tukey.

In the following sections we shall summarize Huber's results.

III.2 Point Estimates Based on Minimal Principle: The M-Estimates

The method of least squares proposes the value which minimizes the sum of squares of differences between observed and expected values as an estimate of the unknown parameter. In the case of estimating a single location parameter one has to minimize the expression $\sum_{i=1}^{n} (X_i - T)^2$. This is achieved by sample mean $T = \frac{1}{n} \sum_{i=1}^{n} X_i$.

It is natural to ask whether one can obtain "more robustness", by minimizing another function of the differences between observed and expected values, than the sum of their squares. Let $X_1, \ldots, X_n$ be a random sample. The estimate $T_n(X) = T_n(X_1, \ldots, X_n)$ of location, based on minimal principle (M-estimate for short) is defined as a solution of the expression

$$(2.1) \qquad \sum_{i=1}^{n} \rho(X_i - T_n(X)) = \text{mimimum} \quad ,$$

where ρ is a non-constant function.

This class of estimates contains in particular

(i) sample mean if $\rho(t) = t^2$

(ii) sample median if $\rho(t) = |t|$.

After defining the class of M-estimates, a criterion of robustness of an estimate must be agreed on, which will allow us to choose the best estimate from a set of estimates (with respect to that criterion). Unfortunately there is no unanimity about this question. We shall accept as a measure of robustness of an estimate, the inverse of the surpremum of the asymptotic variance, when F ranges over a suitable set of underlying distributions in particular over the set of all $F = (1-\varepsilon)\Phi+\varepsilon H$ for fixed ε and H-symmetric. It will be shown, that if we accept this measure of robustness and we restrict attention to M-estimates, then the most robust estimate of location corresponds to

$$\rho(t) = \frac{1}{2} t^2 \qquad \text{for} \qquad |t| < k$$

and

$$\rho(t) = k|t| - \frac{1}{2} k^2 \qquad \text{for} \qquad |t| \geq k ,$$

where k depends on ε.

Thus the most robust estimate of location T_n is defined by

$$\sum_{i=1}^{n} \psi(X_i - T_n) = 0 \quad , \tag{2.2}$$

where $\psi(t) = t$ for $|t| < k$ and $\psi(t) = k.\ \mathrm{sgn}\ (t)$ for $|t| \geq k$ with k depending on ε.

III.3 Asymptotic Normality of M-Estimates for Convex ρ

In this section we are assuming throughout, that $\rho(t)$ is a convex real-valued function of a real variable t, tending to $+\infty$ as $t \to \pm\infty$.

Definition 1: Let $X_1,\ldots,X_n$ be independent, identically distributed random variables, with common distribution function F. Let $[T_n(X)]$ be the set of all those ξ, for which $Q(\xi) = \sum_{i=1}^{n} \rho(X_i-\xi)$ reaches its infimum Q_{inf}. Then we define the M-estimate $T_n(X)$ as any representation of the set valued function $(X_1,\ldots,X_n) \to [T_n(X)]$ by a single valued function $(X_1,\ldots,X_n) \to T_n(X) \in [T_n(X)]$ for instance $T_n(X)$ = midpoint of $[T_n(X)]$ if $[T_n(X)]$ is an interval.

The set $[T_n(X)]$ is invariant under translation, i.e., $[T_n(X+c)] = [T_n(X)]+c$.

LEMMA 1: $Q(\xi)$ is convex function of ξ, and $[T_n(X)]$ is non-empty, convex and compact. If ρ is strictly convex, then $[T_n(X)]$ is reduced to a single point.

PROOF: (Strict) convexity of Q follows from (strict) convexity of ρ.

The sets $E_m = \{\xi \,|\, Q(\xi) \leqq Q_{inf} + \frac{1}{m}\}$, $m = 1,2,\ldots,$ form a decreasing sequence of non-empty, convex, compact sets as $m \to \infty$, hence their intersection $[T_n(X)]$ is non-empty, convex, compact. If ρ is strictly convex, and if ξ', ξ'' were two distinct points of $[T_n(X)]$, then we would have

$$Q_{inf} = Q[\tfrac{1}{2}\xi' + \tfrac{1}{2}\xi''] < \tfrac{1}{2}Q(\xi') + \tfrac{1}{2}Q(\xi'') = Q_{inf} ,$$

which is a contradiction. Hence $\xi' = \xi'' = T_n(X)$.

Let $\psi(t) = \rho'(t)$ be the derivative of $\rho(t)$, normalized such, that $\psi(t) = \frac{1}{2}\psi(t-0) + \frac{1}{2}\psi(t+0)$. ψ is monotone increasing and strictly negative (positive) for large negative (positive) values of t.

<u>Definition 2:</u> If $\psi(t)$ is continuous, then $T_n(X)$ is the solution of the equation

$$\sum_{i=1}^{n} \psi(X_i - T_n(X)) = 0 \quad .$$

<u>Definition 3:</u> Define

$$\lambda(\xi) = \int \psi(t-\xi)dF(t) = E[\psi(X-\xi)] \quad .$$

<u>LEMMA 2:</u> If there is a ξ_o such that $\lambda(\xi_o)$ exists and is finite, then $\lambda(\xi)$ exists for all ξ, (possibly $\lambda(\xi) = \pm\infty$), is montone decreasing and strictly positive (negative) for large negative (positive) values of ξ.

<u>PROOF:</u> Let $\psi = \psi^+ - \psi^-$, where ψ^+ and ψ^- are the positive and negative parts of ψ respectively. Then

$$\lambda(\xi) = \int \psi^+(t-\xi)dF(t) - \int \psi^-(t-\xi)dF(t) \quad .$$

For $\xi=\xi_o$ both integrals exists and are finite. For $\xi \geqq \xi_o$ the first integral is bounded,

$$0 \leqq \int \psi^+(t-\xi)dF(t) \leqq \int \psi^+(t-\xi_o)dF(t) \quad ,$$

and similarly for $\xi \leqq \xi_o$, the second integral is bounded

$$0 \leqq \int \psi^-(t-\xi)dF(t) \leqq \int \psi^-(t-\xi_o)dF(t) \quad .$$

Hence at least one of the two integrals is finite, thus $\lambda(\xi)$ exists everywhere. $\lambda(\xi)$ is monotone decreasing in ξ since $\psi(t-\xi)$ is.

Now we want to show, that $\lambda(\xi)$ is strictly negative for large positive values of ξ and positive for large negative values of ξ. Because of the monotonicity of ψ, it is sufficient to prove the first assertion.

Let $\varepsilon > 0$, and let M be such that

$$\int_M^\infty \psi^+(t-\xi)dF(t) < \varepsilon \; ;$$

for sufficiently large ξ

$$\int_{-\infty}^M \psi^+(t-\xi)dF(t) = 0 \quad ,$$

hence

$$\int_{-\infty}^\infty \psi^+(t-\xi)dF(t) < \varepsilon \quad ,$$

which implies that

$$\int \psi^+(t-\xi)dF(t) \to 0 \quad \text{for} \quad \xi \to \infty \quad .$$

Since ψ takes upon strictly negative values, due to monotonicity of ψ, there is a $\delta > 0$ such that

$$\int \psi^{-}(t-\xi)dF(t) > \delta$$

for sufficiently large ξ, thus $\lambda(\xi)$ is strictly negative for large values of ξ.

LEMMA 3: ("Consistency of T_n.") Assume that there is a c such that $\lambda(\xi) > 0$ for $\xi < c$ and $\lambda(\xi) < 0$ for $\xi > c$. Then $T_n \to c$ almost surely and in probability.

PROOF: Let $\varepsilon > 0$. Then by the law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} \psi(X_i-c-\varepsilon) \to \lambda(c+\varepsilon) = E\psi[X-(c+\varepsilon)]$$

$$\frac{1}{n}\sum_{i=1}^{n} \psi(X_i-c+\varepsilon) \to \lambda(c-\varepsilon) = E\psi[X-(c-\varepsilon)]$$

almost surely and in probability. Hence, by monotonicity of ψ, for almost all sample sequences, $(c-\varepsilon) < [T_n(X)] < (c+\varepsilon)$ holds for some n on, and similarly

$$P\,\{c-\varepsilon < [T_n(X)] < c+\varepsilon\} \to 1 .$$

LEMMA 4: ("Asymptotic normality.") Assume (i) $\lambda(c) = 0$, (ii) $\lambda(\xi)$ is differentiable at $\xi = c$ and $\lambda'(c) < 0$, (iii) $\int\psi^2(t-\xi)dF(t)$ is finite and continuous at $\xi = c$. Then $n^{\frac{1}{2}}(T_n(X)-c)$ is asymptotically normal with asymptotic mean 0 and asymptotic variance

$$V(\psi,F)= \frac{\int \psi^2(t-c)dF(t)}{[\lambda'(c)]^2} .$$

PROOF: Without loss of generality, suppose $c = 0$.

We have to show, that for every fixed real number g,

$$P\{n^{\frac{1}{2}} T_n < g\sigma\} \to \Phi(g) \quad ,$$

where $\sigma = [V(\psi,F)]^{\frac{1}{2}}$. Since from the monotonicity of ψ

$$\{T_n < g\sigma n^{-\frac{1}{2}}\} \subset \{\sum_{i=1}^{n} \psi(X_i - g\sigma n^{-\frac{1}{2}}) < 0\} \subset \{T_n \leqq g\sigma n^{-\frac{1}{2}}\} \quad ,$$

it suffices to show, that

$$p_n = P\{\sum_{i=1}^{n} \psi(X_i - g\sigma n^{-\frac{1}{2}}) < 0\} \to \Phi(g) \quad .$$

Let

$$s^2 = \int [\psi(t - g\sigma n^{-\frac{1}{2}}) - \lambda(g\sigma n^{-\frac{1}{2}})]^2 dF(t) \quad ,$$

then the

$$Y_i = \frac{[\psi(X_i - g\sigma n^{-\frac{1}{2}}) - \lambda(g\sigma n^{-\frac{1}{2}})]}{s}$$

are independent random variables with mean zero and unit variance.

We have

$$p_n = P\{n^{-\frac{1}{2}} \sum_{i=1}^{n} Y_i < -\frac{n^{\frac{1}{2}} \lambda(g\sigma n^{-\frac{1}{2}})}{s}\}$$

and

$$-\frac{n^{\frac{1}{2}} \lambda(g\sigma n^{-\frac{1}{2}})}{s} \to g \quad ,$$

which follows from the assumptions of the lemma, upon applying the

mean value theorem.

We shall see, that $n^{-\frac{1}{2}} \sum_{i=1}^{n} Y_i$ is asymptotically normal with mean 0 and variance 1, hence $p_n \to \Phi(g)$. The

$$Y_i = \frac{1}{s}\,[\psi(X_i - g\sigma n^{-\frac{1}{2}}) - \lambda(g\sigma n^{-\frac{1}{2}})]$$

are independent, identically distributed random variables, but they are different for different values of n, therefore, the normal convergence criterion, as given in Loève [11], (1960), p. 295 will be applied.

The criterion states, that the distribution of $n^{-\frac{1}{2}} \sum_{i=1}^{n} Y_i$ converges toward the standard normal, iff for every $\varepsilon > 0$, as $n \to \infty$,

$$\int_{E} y_1^2 \, dF(x_1) \to 0 \quad ,$$

the integration being extended over the set $E = \{|y_1| \geq n^{\frac{1}{2}}\varepsilon\}$. Since $\lambda(g\sigma n^{-\frac{1}{2}}) \to 0$, this holds iff for every $\varepsilon > 0$

$$\int_{E'} \psi^2(x_1 - n^{-\frac{1}{2}} g\sigma)\, dF(x_1) \to 0 \quad ,$$

the itegration being extended over the set $E' = \{|\psi(x_1 - n^{-\frac{1}{2}} g\sigma)| \geq n^{\frac{1}{2}}\varepsilon\}$.

Now, let $\eta > 0$ be given and let n_o be such that

$$|g\sigma n^{-\frac{1}{2}}| < \eta \qquad \text{for} \qquad n \geq n_o \quad .$$

Then, since ψ is monotone we have

$$\psi^2(x_1 - g\sigma n^{-\frac{1}{2}}) \leqq u^2(x_1) \qquad \text{for} \qquad n \geq n_o \quad ,$$

with $u(x_1) = \max\,\{|\psi(x_1 - \eta)|,\ |\psi(x_1 + \eta)|\}$. Since $E\,\psi^2(x_1 \pm \eta)$ exists

we have

$$\int_{E''} u^2(x_1)dF(x_1) \to 0 \quad ,$$

the integration being extended over the set $E'' = \{|u(x_1)| \geq n^{\frac{1}{2}}\varepsilon\} \supset E'$. This proves the theorem.

For the remainder of this section, let $c = 0$. If ψ and F are sufficiently regular, then

$$\lambda'(0) = [\frac{d}{d\xi} \int \psi(t-\xi)dF(t)]_{\xi=o} =$$

$$= - \int \psi'(t)dF(t) = -E[\psi'] \quad .$$

Moreover, if $F(t)$ is absolute continuous, with density $f(t)$, we find by partial integration

$$\lambda'(0) = \int \psi(t)f'(t)dt \quad .$$

The Schwarz inequality then yields

$$(3.1) \qquad V(\psi,F) = \frac{E\psi^2}{[E\psi']^2} = \frac{\int\psi^2 f dt}{[\int \psi(\frac{f'}{f})fdt]^2} \geq \frac{1}{\int (\frac{f'}{f})^2 fdt} \quad .$$

We have strict inequality unless $\psi = - p \frac{f'}{f}$, for some constant p, that is, unless

$$f(t) = \text{constant} \quad e^{-\frac{\rho(t)}{p}} \quad ,$$

and then the M-estimate is the maximum likelihood estimate.

III.4 Minimax Approach

Asymptotic minimax theory of robust estimation in the case of bounded $\psi(\psi = \rho')$ can be justified from the fact that frequently the sample size is perhaps large enough to indicate deviations from the assumed model, but not yet large enough to establish their nature.

In the case of contaminated normal distribution $F = (1-\varepsilon)\Phi+\varepsilon H$, this means, that the asymptotic minimax theory would be appropriate whenever the sample size is fairly large, but $\varepsilon.n$, the average number of outliers is still rather small.

We shall treat a special case and solve it by a direct verification of saddlepoint property.

Let C be a set of all distributions of the form $F = (1-\varepsilon)G+\varepsilon H$, where $0 \leqq \varepsilon < 1$ is a fixed number, G is a fixed and H a variable distribution function.

Assume, that G has a convex support and a twice continuously differentiable density g, such that $-\log (g)$ is convex on the support of G.

Let T_n be an M-estimate corresponding to a certain ρ, let $\psi = \rho'$ be the derivative of ρ and let c be such that

$$\int \psi(t-c)\,dF(t) = 0 \quad .$$

The asymptotic variance of $n^{\frac{1}{2}}(T_n-c)$ will be

$$V(\psi,F) = \frac{E_F\, \psi^2(t-c)}{[E_F\, \psi'(t-c)]^2} \quad ,$$

provided ψ is sufficiently regular.

We shall try to minimize the supremum $\sup_F V(\psi,F)$ of the asymptotic variance only for those pairs (ψ,F), for which $c = 0$.

THEOREM 1: The asymptotic variance $F(\psi,F)$ has a saddlepoint: there is an $F_o = (1-\varepsilon)G+\varepsilon H_o$ and a ψ_o such that

$$\sup_F V(\psi_o,F) = V(\psi_o,F_o) = \inf_\psi V(\psi,F_o) \quad ,$$

where F ranges over those distributions in C, for which $E_F\, \psi_o = 0$.

Let $t_o < t_1$, be the endpoints of the interval, where $|\frac{g'}{g}| \leqq k$ (either or both of these endpoints may be at infinity), and k is related to ε by

$$(1-\varepsilon)^{-1} = \int_{t_o}^{t_1} g(t)dt = \frac{[g(t_o)+g(t_1)]}{k} \quad .$$

Then the density f_o of F_o is given by

$$\begin{aligned} f_o(t) &= (1-\varepsilon)g(t_o)\, e^{k(t-t_o)} && \text{for} \quad t \leq t_o \\ &= (1-\varepsilon)g(t) && \text{for} \quad t_o < t < t_1 \\ &= (1-\varepsilon)g(t_1)\, e^{-k(t-t_1)} && \text{for} \quad t \geq t_1 \quad . \end{aligned}$$

$\psi_o = -\frac{f'_o}{f_o}$ is monotone and bounded and corresponds to a maximum likelihood estimate of the location parameter when F_o is the underlying distribution.

Remark: The statement of this theorem, is unsatisfactory insofar as the class over which H ranges depends on ψ_o. This could be avoided by restricting G to be symmetric, and letting H range over all symmetric distributions.

PROOF: The total mass of F_o is 1, since

$$\int_{-\infty}^{\infty} f_o(t)dt = [\int_{-\infty}^{t_o} f_o(t)dt + \int_{t_o}^{t_1} f_o(dt) + \int_{t_1}^{\infty} f_o(t)dt] =$$

$$= (1-\varepsilon)[g(t_o)\int_{-\infty}^{t_o} e^{k(t-t_o)}dt + \int_{t_o}^{t_1} g(t)dt + \int_{t_1}^{\infty} g(t_1)e^{-k(t-t_1)}dt]$$

$$= (1-\varepsilon)[g(t_o)\frac{1}{k} + \int_{t_o}^{t_1} g(t)dt + g(t_1)\frac{1}{k}] \quad ,$$

and the last expression according to the relation between k and ε is equal to 1.

From this it follows that $H_o = \frac{1}{\varepsilon}[F_o-(1-\varepsilon)G]$ has a total mass 1. It remains to check, that h_o is non-negative. But

$$\begin{aligned} \varepsilon h_o(t) &= (1-\varepsilon)[g(t_o)e^{k(t-t_o)}-g(t)] && \text{for} \quad t \leq t_o \\ &= 0 && \text{for} \quad t_o < t < t_1 \\ &= (1-\varepsilon)[g(t_1)e^{-k(t-t_1)}-g(t)] && \text{for} \quad t \geq t_1 \quad . \end{aligned}$$

Because the function $-\log g(t)$ is convex, it lies above its tangents at t_o and t_1, i.e., $-\log g(t) \geq -\log g(t_o) - k(t-t_o)$, thus $g(t) \leqq g(t_o)e^{k(t-t_o)}$. Similarly for the point t_1, we have $g(t) \leqq g(t)e^{-k(t-t_1)}$. This implies the non-negativity of $h_o(t)$.

It is easy to see that $\psi_o(t) = -\frac{f_o'(t)}{f_o(t)}$ is bounded and monotone, so we have

$$V(\psi_o,F_o) = \frac{E_{F_o}\ \psi_o^2}{[E_{F_o}\ \psi_o']^2} = \frac{(1-\varepsilon)\ E_G\ \psi_o^2 + \varepsilon k^2}{[(1-\varepsilon)\ E_G\ \psi_o']^2}\ .$$

The right side is an obvious upper bound for $V(\psi_o,F)$ provided $E_F\ \psi_o = 0$, so we have $V(\psi_o,F) \leqq V(\psi_o,F_o)$ for all $F \in C$.

The inequality $V(\psi_o,F_o) \leqq V(\psi,F_o)$ follows directly from inequality (3.1) of the previous section, noticing that

$$V(\psi_o,F_o) = \frac{1}{\int (\frac{f_o'}{f_o})^2\ f_o dt}$$

This proves the theorem.

It can be shown, using results of Le Cam [9], [10], (1953), (1958), that the (M)-estimate, corresponding to ψ_o, minimizes the maximal asymptotic variance not only among (M)-estimates, but even among all translation invariant estimates.

The assumptions of Theorem 1 are satisfied, if $G = \Phi$ is the standard normal distribution, with density $\phi(t) = (2\pi)^{-\frac{1}{2}} \exp\{-\frac{1}{2}t^2\}$ i.e., in the case of contaminated normal distribution. We summarize this in

<u>Corollary:</u> Define the estimate $T_n = T_n(X_1,\ldots,X_n)$ by the property, that it minimizes

$$\sum_{i=1}^{n} \rho(X_i - T_n)\quad ,$$

where

$$\rho(t) = \frac{1}{2} t^2 \qquad \text{for} \quad |t| < k$$
$$= k\,|t| - \frac{1}{2} k^2 \qquad \text{for} \quad |t| \geq k \,.$$

Let $\psi = \rho'$. For symmetric H, $n^{\frac{1}{2}} T_n$ is asymptotically normal with asymptotic mean 0, and attains the maximal asymptotic variance

$$\sup_F \; V(\psi,F) = \frac{\{(1-\varepsilon)E_\Phi \; \psi^2+\varepsilon k^2\}}{\{(1-\varepsilon)\; E_\Phi \; \psi'\}^2} \quad ,$$

whenever H puts all its mass outside the interval $[-k,k]$. If ε and k are related by

$$(1-\varepsilon)^{-1} = \int_{-k}^{k} \phi(t)dt + \frac{2}{k}\,\phi(k) \quad ,$$

then T_n minimizes the maximal asymptotic variance among all translation invariant estimates, the maximum being taken over the set of all symmetric ε-contaminated distributions $F = (1-\varepsilon)\Phi+\varepsilon H$. There is a unique asymptotically least favourable F_o, having the density

$$f_o(t) = (1-\varepsilon)\,\frac{1}{\sqrt{2\pi}}\, e^{-\rho(t)} \quad .$$

CHAPTER IV

INTERMEDIARIES BETWEEN SAMPLE MEAN AND SAMPLE MEDIAN

IV.1 Introduction

In this chapter some new estimates of location which, in a sense, lie between sample mean and median will be presented.

For the one-sample problem of estimating a single location parameter μ $(-\infty < \mu < \infty)$ of a continuous and symmetric distribution $F(x-\mu)$ i.e., the center of symmetry of the underlying distribution $F(x-\mu)$, we accept the model of indeterminacy, with the prototype distribution of the form:

$$F_{\varepsilon,h}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(\frac{x}{h}) \quad ,$$

where $\Phi(x)$ is the standard normal distribution, and, ε is a fixed fraction of contamination $0 \leqq \varepsilon \leqq \frac{1}{2}$, and h is the scale ratio of the two component normal distributions $1 < h \leqq h^* < \infty$, with h^* fixed.

We shall confine our attention to the class of estimates of the unknown center of symmetry μ, which are intermediaries between the sample mean and sample median in the sense, that the sample mean and sample median will be limiting cases in the class of estimates. To determine a most robust estimate in the class of estimates, we will accept as a measure of robustness of an estimate the inverse of the supremum of the asymptotic variance, when $F_{\varepsilon,h}(x-\mu)$ ranges over the set

$$(1.1) \qquad \mathcal{F} = \{F_{\varepsilon,h}(x-\mu) : F_{\varepsilon,h}(x-\mu) = (1-\varepsilon)\Phi(x-\mu)+\varepsilon\Phi(\frac{x-\mu}{h}) \;, \quad 1 < h \leqq h^* < \infty\} \;,$$

for ε and h^* fixed. In a practical situation the class $\mathcal{F}$ would be relevant if one can assume that the variance of the contaminating distribution ("wider" normal distribution) has an upper bound.

The reason for confining our attention to intermediaries between sample mean and sample median is the following.

The concept of a robust estimate even in the simplest case of the one-sample problem of estimating a single location parameter of a symmetric and continuous distribution permits different interpretations. Intuitively it corresponds to an estimate, (or rather to a sequence of estimates), which (i) can tolerate at least a few outliers, i.e., which can tolerate at least a few grossly erroneous observations in the sample, but at the same time it (ii) has small asymptotic variance when there is no contamination present. Since in our accepted model of indeterminacy the sample mean satisfies (ii) but not (i) and the sample median satisfies (i) but not (ii) we tried for some compromise.

IV.2 Estimates for the One-Sample Location Problem.

In this section we shall introduce two classes of estimates, in which the estimates are intermediaries between sample mean and sample median (in the above sense), and which are based on minimal principle.

Definition 1: Let

$$(2.1) \qquad \rho_1(t;\lambda) = |t|^{1+\lambda} \quad \text{for} \quad -\infty < t < \infty \quad \text{and} \quad 0 \leq \lambda \leq 1 .$$

For each fixed $0 \leq \lambda \leq 1$ we define the estimate T_λ^n corresponding to the function $\rho_1(t;\lambda)$ by

$$(2.2) \qquad \sum_{i=1}^{n} \rho_1(X_i - T_\lambda^n) = \text{minimum} .$$

Definition 2: Let

$$(2.3) \qquad \rho_2(t;\nu) = (1-\nu)|t| + \nu t^2 \quad \text{for} \quad -\infty < t < \infty \quad \text{and} \quad 0 \leq \nu \leq 1 .$$

For each fixed $0 \leq \nu \leq 1$ we define the estimate T_ν^n corresponding to the function $\rho_2(t;\nu)$ by

$$(2.4) \qquad \sum_{i=1}^{n} \rho_2(X_i - T_\nu^n) = \text{minimum} .$$

The functions ρ_1 and ρ_2 both satisfy the conditions of Huber's ρ-function, i.e., for each fixed $0 \leq \lambda , \nu \leq 1$ they are continuous, convex real functions of one real variable t. Hence the estimates T_λ^n and T_ν^n of Definitions 1 and 2 are based on minimal principle, i.e., they are M-estimates for each fixed $0 \leq \lambda , \nu \leq 1$. Further, for $\lambda = \nu = 0$ we have $\rho_1(t) = \rho_2(t) = |t|$ and for $\lambda = \nu = 1$ we have $\rho_1(t) = \rho_2(t) = t^2$, and therefore the estimates corresponding to these limiting cases are the sample median and sample mean, respectively, since (2.2) and (2.4) then lead

$$\sum_{i=1}^{n} |X_i - T^n_{\lambda=o}| = \text{minimum} \iff T^n_{\lambda=o} = \text{sample median}$$

$$\sum_{i=1}^{n} |X_i - T^n_{\nu=o}| = \text{minimum} \iff T^n_{\nu=o} = \text{sample median}$$

and

$$\sum_{i=1}^{n} (X_i - T^n_{\lambda=1})^2 = \text{minimum} \iff T^n_{\lambda=1} = \text{sample mean}$$

$$\sum_{i=1}^{n} (X_i - T^n_{\nu=1})^2 = \text{minimum} \iff T^n_{\nu=1} = \text{sample mean} \quad .$$

The prototype distribution function - the contaminated normal distribution

(2.5) $$F_{\varepsilon,h}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(\frac{x}{h}) \quad ,$$

where $\Phi(x)$ is the standard normal distribution, ε is a fixed fraction of contamination and h is scale ratio of the component normal distributions, $1 < h \leqq h^* < \infty$, with h^* fixed, will be used, for the problem of estimating the unknown center of symmetry μ of the distribution $F_{\varepsilon,h}(x-\mu)$.

Let $X_1,\ldots,X_n$ be independent, identically distributed random variables with common distribution function $F_{\varepsilon,h}(x-\mu)$, where the prototype distribution $F_{\varepsilon,h}(x)$ is defined by (2.5).

Let

(2.6) $$E_1 = \{T^n_\lambda = T^n_\lambda(X_1,\ldots,X_n) : \sum_{i=1}^{n} \rho_1(X_i - T^n_\lambda)$$

$$= \sum_{i=1}^{n} |X_i - T^n_\lambda|^{1+\lambda} = \min, \quad 0 \leqq \lambda \leqq 1\} \quad ,$$

and

$$(2.7) \qquad E_2 = \{T_\nu^n = T_\nu^n(X_1,\ldots,X_n): \sum_{i=1}^n \rho_2(X_i - T_\nu^n) = (1-\nu)\sum_{i=1}^n |X_i - T_\nu^n| + \nu \sum_{i=1}^n (X_i - T_\nu^n)^2 = \min, \quad 0 \leqq \nu \leqq 1\} ,$$

be two classes of estimates corresponding to functions $\rho_1(t;\lambda)$ and $\rho_2(t;\nu)$, respectively.

If $\psi_1(t;\lambda) = \rho_1'(t;\lambda)$ and $\psi_2(t;\nu) = \rho_2'(t;\nu)$ are the normalized derivatives $\psi(t) = \frac{1}{2}\psi(t-0) + \frac{1}{2}\psi(t+0)$, then

$$\psi_1(t;\lambda) = (1+\lambda)|t|^\lambda \operatorname{sgn}(t) , \qquad \psi_2(t;\nu) = (1-\nu)\operatorname{sgn}(t) + 2\nu t ,$$

and the class E_1 can be equivalently defined by

$$(2.8) \qquad E_1 = \{T_\lambda^n: \sum_{i=1}^n \psi_1(X_i - T_\lambda^n) = (1+\lambda)\sum_{i=1}^n |X_i - T_\lambda^n|^\lambda \operatorname{sgn}(X_i - T_\lambda^n) = 0 , \quad 0 \leqq \lambda \leqq 1\} .$$

We will determine according to our measure of robustness the most robust estimates of the unknown location parameter μ of the distribution $F_{\varepsilon,h}(x-\mu)$, in the two classes E_1 and E_2, i.e., to determine the best choice of parameters λ and ν for that problem.

IV.3 Regularity Properties of the Estimates T_λ^n and T_ν^n.

Let us define

$$(3.1) \qquad \Lambda_1(\xi) = \int_{-\infty}^{\infty} \psi_1(t-\xi)\, dF_{\varepsilon,h}(t-\mu) ,$$

and

$$\Lambda_2(\xi) = \int_{-\infty}^{\infty} \psi_2(t-\xi)\,dF_{\varepsilon,h}(t-\mu) \quad . \tag{3.2}$$

We have

$$\Lambda_1(\xi) = \int_{-\infty}^{\infty} (1+\lambda)|t-\xi|^{\lambda} \operatorname{sgn}(t-\xi) f_{\varepsilon,h}(t-\mu)\,dt =$$

$$= \int_{-\infty}^{\infty} [(1+\lambda)|t-\xi|^{\lambda} \operatorname{sgn}(t-\xi)][\frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\mu)^2} + \frac{\varepsilon}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\mu}{h})^2}]\,dt \quad ,$$

and

$$\Lambda_2(\xi) = \int_{-\infty}^{\infty} [(1-\nu) \operatorname{sgn}(t-\xi)+2\nu(t-\xi)] f_{\varepsilon,h}(t-\mu)\,dt =$$

$$= \int_{-\infty}^{\infty} [(1-\nu) \operatorname{sgn}(t-\xi)+2\nu(t-\xi)][\frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\mu)^2} + \frac{\varepsilon}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\mu}{h})^2}]dt \quad .$$

It is easy to see, that for $\xi = \mu$, both integrals $\Lambda_1(\xi)$ and $\Lambda_2(\xi)$ exist, and are equal to zero, since

$$\Lambda_1(\mu) = \int_{-\infty}^{\infty} (1+\lambda)|t-\mu|^{\lambda} \operatorname{sgn}(t-\mu)[\frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\mu)^2} + \frac{\varepsilon}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\mu}{h})^2}]dt$$

$$= \int_{o}^{\infty} (1+\lambda)z^{\lambda}[\frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} + \frac{\varepsilon}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{z}{h})^2}]dz -$$

$$- \int_{o}^{\infty} (1+\lambda)z^{\lambda}\,[\frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} + \frac{\varepsilon}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{z}{h})^2}]dz = 0 \quad ,$$

and

$$\Lambda_2(\mu) = \int_{-\infty}^{\infty} [(1-\nu)\text{sgn}(t-\mu)+2\nu(t-\mu)][\frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\mu)^2} + \frac{\varepsilon}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\mu}{h})^2}]dt$$

$$= (1-\nu)[\int_{o}^{\infty} f_{\varepsilon,h}(t-\mu)dt - \int_{-\infty}^{o} f_{\varepsilon,h}(t-\mu)dt]$$

$$+ 2\nu \int_{-\infty}^{\infty} (t-\mu)[\frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\mu)^2} + \frac{\varepsilon}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\mu}{h})^2}]dt$$

$$= (1-\nu).0 + 2\nu.0 = 0 \quad .$$

Then from Lemma 2 of Chapter 3 it follows, that $\Lambda_1(\xi)$ and $\Lambda_2(\xi)$ exists for all ξ, and both are monotone decreasing functions of ξ.

Since for $\xi = \mu$ we have $\Lambda_1(\mu) = \Lambda_2(\mu) = 0$, we can apply also Lemma 3 of Chapter 3, from which it follows that

$$T_\lambda^n(X_1,\ldots,X_n) \to \mu$$

almost surely and in probability for all $0 \le \lambda \le 1$, and

$$T_\nu^n(X_1,\ldots,X_n) \to \mu$$

almost surely and in probability for all $0 \le \nu \le 1$.

In order to show the asymptotic normality and to find the asymptotic variance of $T_\lambda^n \in E_1$ and $T_\nu^n \in E_2$, we shall use the result of Lemma 4 of Chapter 3.

We have to show, that the assumptions of Lemma 4 are satisfied for both $T_\lambda^n \in E_1$ and $T_\nu^n \in E_2$. The assumptions we have to verify are:

(i) $\Lambda_j(\mu) = 0$

(ii) $\Lambda_j(\xi)$ is differentiable at $\xi = \mu$ and $\Lambda_j'(\mu) < 0$

(iii) $I_j(\xi) = \int_{-\infty}^{\infty} \psi_j^2(t-\xi)dF_{\varepsilon,h}(t-\mu)$ is finite and continuous at $\xi = \mu$; for $j = 1,2$.

Assumption (i) is satisfied for $j = 1,2$, since we already showed that $\Lambda_1(\mu) = \Lambda_2(\mu) = 0$. Let us show first that the assumptions (ii) and (iii) are satisfied for $j = 1$. From (3.1) we have

$$\Lambda_1(\xi) = \int_{-\infty}^{\infty} \psi_1(t-\xi)dF_{\varepsilon,h}(t-\mu) \quad ,$$

where

$$\psi_1(t-\xi) = (1+\lambda)|t-\xi|^{\lambda} \operatorname{sgn}(t-\xi) \ ,$$

$$-\infty < t < \infty, \quad -\infty < \xi < \infty, \quad 0 \leq \lambda \leq 1 \ .$$

To prove that $\Lambda_1(\xi)$ is differentiable at $\xi = \mu$ we shall split $\Lambda_1(\xi)$ into four integrals and check the conditions for differentiability under the integral sign (see for example, Cramér [3], p. 67) for each of them.

The integral $\Lambda_1(\xi)$ can be written in the following form:

$$\Lambda_1(\xi) = \int_{-\infty}^{\infty} [(1+\lambda)|t-\xi|^{\lambda}\operatorname{sgn}(t-\xi)][\frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\mu)^2} + \frac{\varepsilon}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\mu}{h})^2}]dt$$

$$= \int_{-\infty}^{\infty} [(1+\lambda)|t-\xi|^{\lambda}\operatorname{sgn}(t-\xi)] \frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\mu)^2} dt$$

$$+ \int_{-\infty}^{\infty} [(1+\lambda)|t-\xi|^{\lambda}\operatorname{sgn}(t-\xi)] \frac{\varepsilon}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\mu}{h})^2} dt =$$

$$= \frac{(1+\lambda)(1-\varepsilon)}{\sqrt{2\pi}} \{\int_0^\infty t^\lambda e^{-\frac{1}{2}[t+(\xi-\mu)]^2} dt - \int_0^\infty t^\lambda e^{-\frac{1}{2}[t-(\xi-\mu)]^2} dt\}$$

$$+ \frac{(1+\lambda)\varepsilon h^\lambda}{\sqrt{2\pi}} \{\int_0^\infty t^\lambda e^{-\frac{1}{2}[t+(\frac{\xi-\mu}{h})]^2} dt - \int_0^\infty t^\lambda e^{-\frac{1}{2}[t-(\frac{\xi-\mu}{h})]^2} dt\}$$

$$= \frac{(1+\lambda)(1-\varepsilon)}{\sqrt{2\pi}} \{\int_0^\infty g_1(t,\xi)dt - \int_0^\infty g_2(t,\xi)dt\}$$

$$+ \frac{(1+\lambda)\varepsilon h^\lambda}{\sqrt{2\pi}} \{ \int_0^\infty g_3(t,\xi)dt - \int_0^\infty g_4(t,\xi)dt\}, \quad \text{say}$$

where $0 \leq \lambda \leq 1$, $0 \leq \varepsilon \leq \frac{1}{2}$, $\xi \in (-\infty,\infty)$ and $1 < h \leq h^* < \infty$ with h^* fixed.

(a) It is easy to see, that the functions $g_i(t,\xi)$, $i = 1,2,3,4$ are for each fixed $\xi \in (-\infty,\infty)$ Borel measurable functions in t, since the sets $A_i = \{t:g_i(t,\xi) \leq k\}$ for $i = 1,2,3,4$ are Borel sets, for each real k.

(b) For each fixed $\xi \in (-\infty,\infty)$ the functions $g_i(t,\xi)$, $i = 1,2,3,4$ are integrable functions, since for instance

$$\int_0^\infty |g_1(t,\xi)|dt = \int_0^\infty t^\lambda e^{-\frac{1}{2}[t+(\xi-\mu)]^2} dt$$

$$= \int_0^1 t^\lambda e^{-\frac{1}{2}[t+(\xi-\mu)]^2} dt + \int_1^\infty t^\lambda e^{-\frac{1}{2}[t+(\xi-\mu)]^2} dt$$

$$\leq 1 + \int_{1+(\xi-\mu)}^\infty [y-(\xi-\mu)]\, e^{-\frac{1}{2}y^2} dy$$

$$= 1 + e^{-\frac{1}{2}[1+(\xi-\mu)]^2} - (\xi-\mu)\sqrt{2\pi}\,[1-\Phi[1+(\xi-\mu)]]$$

$$\leq 2 + |\xi-\mu| \sqrt{2\pi} < \infty$$

where $\Phi(x)$ is the standard normal distribution. The integrability of functions $g_i(t,\xi)$, for $i = 2,3,4$ can be shown similarly.

(c) The partial derivatives $\frac{\partial}{\partial\xi} g_i(t,\xi)$, $i = 1,2,3,4$ exists for each fixed $\xi \in (-\infty,\infty)$ and for all $t \in (0,\infty)$, since

$$\frac{\partial}{\partial\xi} g_1(t,\xi) = \frac{\partial}{\partial\xi} \{t^\lambda e^{-\frac{1}{2}[t+(\xi-\mu)]^2}\} = -t^\lambda [t+(\xi-\mu)]\, e^{-\frac{1}{2}[t+(\xi-\mu)]^2}$$

$$\frac{\partial}{\partial\xi} g_2(t,\xi) = \frac{\partial}{\partial\xi} \{t^\lambda e^{-\frac{1}{2}[t-(\xi-\mu)]^2}\} = t^\lambda [t-(\xi-\mu)]\, e^{-\frac{1}{2}[t-(\xi-\mu)]^2}$$

$$\frac{\partial}{\partial\xi} g_3(t,\xi) = \frac{\partial}{\partial\xi} \{t^\lambda e^{-\frac{1}{2}[t+(\frac{\xi-\mu}{h})]^2}\} = -\frac{1}{h}\, t^\lambda [t+(\frac{\xi-\mu}{h})]\, e^{-\frac{1}{2}[t+(\frac{\xi-\mu}{h})]^2}$$

$$\frac{\partial}{\partial\xi} g_4(t,\xi) = \frac{\partial}{\partial\xi} \{t^\lambda e^{-\frac{1}{2}[t-(\frac{\xi-\mu}{h})]^2}\} = \frac{1}{h}\, t^\lambda [t-(\frac{\xi-\mu}{h})]\, e^{-\frac{1}{2}[t-(\frac{\xi-\mu}{h})]^2} .$$

(d) The absolute values of partial derivatives $\frac{\partial}{\partial\xi} g_i(t,\xi)$, $i = 1,2,3,4$ are bounded by a function $G_1(t)$, defined by:

$$\begin{aligned} G_1(t) &= (t+1)\,t\, e^{-\frac{1}{2}(t-1)^2} && \text{for} \quad t \geq 1 \\ &= (t+1) && \text{for} \quad 0 < t < 1 . \end{aligned}$$

We have for $i = 1,2$ and for all $\xi \in (\mu-1,\ \mu+1)$,

$$\left|\frac{\partial}{\partial\xi} g_i(t,\xi)\right| < G_1(t)$$

and for $i = 3,4$ and for all $\xi \in (\mu-h,\ \mu+h)$

$$\left|\frac{\partial}{\partial\xi} g_i(t,\xi)\right| < G_1(t) .$$

The function $G_1(t)$ is integrable since

$$\int_0^\infty G_1(t)dt = \int_0^1 (t+1)dt + \int_1^\infty (t+1)t\, e^{-\frac{1}{2}(t-1)^2} dt$$

$$= \frac{1}{2}[4-1] + \int_0^\infty [y^2+3y+2]\, e^{-\frac{1}{2}y^2} dy$$

$$= \frac{3}{2} + [3 + \sqrt{2\pi} + \frac{\sqrt{2\pi}}{2}] = 3 + \frac{3}{2}[1 + \sqrt{2\pi}] < \infty$$

Hence the usual conditions for differentiation under the integral sign are satisfied and we have

$$(3.3) \qquad A_1 = \frac{d}{d\xi} \Lambda_1(\xi)\Big|_{\xi=\mu} =$$

$$= \frac{(1+\lambda)(1-\varepsilon)}{\sqrt{2\pi}} \{\int_0^\infty \frac{\partial}{\partial\xi} g_1(t,\xi)\Big|_{\xi=\mu} dt - \int_0^\infty \frac{\partial}{\partial\xi} g_2(t,\xi)\Big|_{\mu=\xi} dt\}$$

$$+ \frac{(1+\lambda)\varepsilon h^\lambda}{\sqrt{2\pi}} \{\int_0^\infty \frac{\partial}{\partial\xi} g_3(t,\xi)\Big|_{\mu=\xi} dt - \int_0^\infty \frac{\partial}{\partial\xi} g_4(t,\xi)\Big|_{\mu=\xi} dt\}$$

$$= \frac{(1+\lambda)(1-\varepsilon)}{\sqrt{2\pi}} \{[-\int_0^\infty t^{\lambda+1} e^{-\frac{1}{2}t^2} dt] - [\int_0^\infty t^{\lambda+1} e^{-\frac{1}{2}t^2} dt]\}$$

$$+ \frac{(1+\lambda)\varepsilon h^\lambda}{\sqrt{2\pi}} \{[-\frac{1}{h}\int_0^\infty t^{\lambda+1} e^{-\frac{1}{2}t^2} dt] - [\frac{1}{h}\int_0^\infty t^{\lambda+1} e^{-\frac{1}{2}t^2} dt]\}$$

$$= -\frac{(1+\lambda).2}{\sqrt{2\pi}} [\int_0^\infty t^{\lambda+1} e^{-\frac{1}{2}t^2} dt]\{(1-\varepsilon) + \varepsilon h^{\lambda-1}\}$$

$$= -\frac{(1+\lambda).2}{\sqrt{2\pi}} [2^{\frac{\lambda}{2}} \Gamma(\frac{\lambda}{2}+1)]\{(1-\varepsilon) + \varepsilon h^{\lambda-1}\}$$

$$= -\frac{(1+\lambda).2^{\frac{\lambda}{2}+1}}{\sqrt{2\pi}} \Gamma(\frac{\lambda}{2}+1)\ \{(1-\varepsilon) + \varepsilon h^{\lambda-1}\} < 0\ .$$

This shows that $\Lambda_1(\xi)$ is differentiable at $\xi = \mu$. Since $A_1 = \Lambda_1'(\mu) < 0$, we have that requirement (ii) of Lemma 4 is satisfied for $j = 1$.

Next we want to show, that the integral

$$I_1(\xi) = \int_{-\infty}^{\infty} \psi_1^2(t-\xi)\,dF_{\varepsilon,h}(t-\mu)$$

is finite and continuous at $\xi = \mu$, where

$$\psi_1^2(t-\xi) = (1+\lambda)^2\ |t-\xi|^{2\lambda}\ [\operatorname{sgn}\,(t-\xi)]^2\ ,$$

$$-\infty < t < \infty, \qquad -\infty < \xi < \infty, \qquad 0 \le \lambda \le 1\ .$$

We shall split $I_1(\xi)$ into four integrals and check the conditions for continuity of the parametric integral (see for example, Cramér [3], p. 67] for each of them.

The integral $I_1(\xi)$ can be written in the following form:

$$I_1(\xi) = \int_{-\infty}^{\infty} [(1+\lambda)^2|t-\xi|^{2\lambda}[\operatorname{sgn}(t-\xi)]^2][\frac{(1-\varepsilon)}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(t-\mu)^2} + \frac{\varepsilon}{h\sqrt{2\pi}}\, e^{-\frac{1}{2}(\frac{t-\mu}{h})^2}\}dt$$

$$= \int_{-\infty}^{\infty} [(1+\lambda)^2|t-\xi|^{2\lambda}\ \frac{(1-\varepsilon)}{\sqrt{2\pi}}\ e^{-\frac{1}{2}(t-\mu)^2}\,dt$$

$$+ \int_{-\infty}^{\infty} [(1+\lambda)^2|t-\xi|^{2\lambda}\ \frac{\varepsilon}{h\sqrt{2\pi}}\ e^{-\frac{1}{2}(\frac{t-\mu}{h})^2}\,dt$$

$$= \frac{(1-\varepsilon)(1+\lambda)^2}{\sqrt{2\pi}}\ \{\int_{o}^{\infty} t^{2\lambda}\ e^{-\frac{1}{2}[t+(\xi-\mu)]^2}\,dt + \int_{o}^{\infty} t^{2\lambda}\ e^{-\frac{1}{2}[t-(\xi-\mu)]^2}\,dt\}$$

$$+ \frac{\varepsilon h^{2\lambda}(1+\lambda)^2}{\sqrt{2\pi}}\ \{\int_{o}^{\infty} t^{2\lambda}\ e^{-\frac{1}{2}[t+(\frac{\xi-\mu}{h})]^2}\,dt + \int_{o}^{\infty} t^{2\lambda}\ e^{-\frac{1}{2}[t-(\frac{\xi-\mu}{h})]^2}\,dt\} =$$

$$= \frac{(1-\varepsilon)(1+\lambda)^2}{\sqrt{2\pi}} \{\int_0^\infty g_5(t,\xi)dt + \int_0^\infty g_6(t,\xi)dt\}$$

$$+ \frac{\varepsilon h^{2\lambda}(1+\lambda)^2}{\sqrt{2\pi}} \{\int_0^\infty g_7(t,\xi)dt + \int_0^\infty g_8(t,\xi)dt\}, \text{ say,}$$

where $0 \leq \lambda \leq 1$, $0 \leq \varepsilon \leq \frac{1}{2}$, $\xi \in (-\infty,\infty)$, and $1 < h \leq h^* < \infty$ with h^* fixed.

(a) It is easy to see that the functions $g_i(t,\xi)$, $i = 5,6,7,8$ are for each fixed $\xi \in (-\infty,\infty)$ Borel measurable functions in t, since the sets $A_i = \{t:g_i(t,\xi) \leq k\}$ for $i = 5,6,7,8$ are Borel sets, for each real k.

(b) For each fixed $\xi \in (-\infty,\infty)$, the functions $g_i(t,\xi)$, $i = 5,6,7,8$ are integrable functions, since for example

$$\int_0^\infty |g_5(t,\xi)|dt = \int_0^\infty t^{2\lambda} e^{-\frac{1}{2}[t+(\xi-\mu)]^2} dt$$

$$= \int_0^1 t^{2\lambda} e^{-\frac{1}{2}[t+(\xi-\mu)]^2} dt + \int_1^\infty t^{2\lambda} e^{-\frac{1}{2}[t+(\xi-\mu)]^2} dt$$

$$\leq 1 + \int_{1+(\xi-\mu)}^\infty [y^2-2(\xi-\mu)y+(\xi-\mu)^2] e^{-\frac{1}{2}y^2} dy$$

$$\leq 1 + \sqrt{2\pi} \int_{1+(\xi-\mu)}^\infty y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy + 2|\xi-\mu| \int_{1+(\xi-\mu)}^\infty y\, e^{-\frac{1}{2}y^2} dy$$

$$+ (\xi-\mu)^2 \sqrt{2\pi} \int_{1+(\xi-\mu)}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

$$\leq 1 + \sqrt{2\pi} + 2|\xi-\mu| + (\xi-\mu)^2 \sqrt{2\pi} \leq 1 + \sqrt{2\pi}\,[1+|\xi-\mu|]^2 < \infty .$$

The integrability of functions $g_i(t,\xi)$ for $i = 6,7,8$ can be shown

similarly.

(c) The functions $g_i(t,\xi)$, $i = 5,6,7,8$ are continuous functions in ξ at $\xi = \mu$ for all $t \in (0,\infty)$, which follows directly from the expressions for the functions $g_i(t,\xi)$, $i = 5,6,7,8$.

(d) The absolute values of the functions $g_i(t,\xi)$, $i = 5,6,7,8$ are bounded by a function $G_2(t)$ defined by:

$$G_2(t) = t^2 e^{-\frac{1}{2}(t-1)^2} \qquad \text{for} \qquad t \geq 1$$

$$= 1 \qquad \text{for} \qquad 0 < t < 1 \quad .$$

We have for $i = 5,6$ and for all $\xi \in (\mu-1, \mu+1)$,

$$|g_i(t,\xi)| < G_2(t)$$

and for $i = 7,8$ and for all $\xi \in (\mu-h, \mu+h)$,

$$|g_i(t,\xi)| < G_2(t) \ .$$

The function $G_2(t)$ is integrable since

$$\int_0^\infty G_2(t)dt = \int_0^1 1.dt + \int_1^\infty t^2 e^{-\frac{1}{2}(t-1)^2} dt$$

$$= 1 + \int_0^\infty (y+1)^2 e^{-\frac{1}{2}y^2} dy$$

$$= 1 + \int_0^\infty (y^2+2y+1) e^{-\frac{1}{2}y^2} dy$$

$$= 1 + \frac{\sqrt{2\pi}}{2}\int_{-\infty}^\infty y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy + 2[-e^{-\frac{1}{2}y^2}]_0^\infty + \frac{\sqrt{2\pi}}{2}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

$$= 1 + \frac{\sqrt{2\pi}}{2} + 2 + \frac{\sqrt{2\pi}}{2} = 3 + \sqrt{2\pi} < \infty \quad ;$$

Hence the usual conditions for continuity of the parametric integral are satisfied and we have

$$(3.4) \qquad B_1 = \lim_{\xi \to \mu} I_1(\xi) = \frac{(1-\varepsilon)(1+\lambda)^2}{\sqrt{2\pi}} \{\int_0^\infty g_5(t,\mu)dt + \int_0^\infty g_6(t,\mu)dt\}$$

$$+ \frac{\varepsilon h^{2\lambda}(1+\lambda)^2}{\sqrt{2\pi}} \{\int_0^\infty g_7(t,\mu)dt + \int_0^\infty g_8(t,\mu)dt\}$$

$$= \frac{(1+\lambda)^2(1-\varepsilon).2}{\sqrt{2\pi}} [\int_0^\infty t^{2\lambda} e^{-\frac{1}{2}t^2} dt]$$

$$+ \frac{\varepsilon h^{2\lambda}(1+\lambda)^2.2}{\sqrt{2\pi}} [\int_0^\infty t^{2\lambda} e^{-\frac{1}{2}t^2} dt]$$

$$= \frac{(1+\lambda)^2.2}{\sqrt{2\pi}} [\int_0^\infty t^{2\lambda} e^{-\frac{1}{2}t^2} dt] \{(1-\varepsilon)+\varepsilon.h^{2\lambda}\}$$

$$= \frac{(1+\lambda)^2.2}{\sqrt{2\pi}} [2^{\lambda-\frac{1}{2}} \Gamma(\lambda+\tfrac{1}{2})] \{(1-\varepsilon)+\varepsilon.h^{2\lambda}\}$$

$$= \frac{(1+\lambda)^2.2^{\lambda+\frac{1}{2}}}{\sqrt{2\pi}} \Gamma(\lambda+\tfrac{1}{2}) \{(1-\varepsilon)+\varepsilon.h^{2\lambda}\} .$$

This shows that $I_1'(\xi)$ is continuous at $\xi = \mu$, and $I_1(\mu)$ is finite for all $0 \leq \varepsilon \leq \frac{1}{2}$, $0 \leq \lambda \leq 1$, $1 < h \leq h^* < \infty$ with h^* fixed, hence requirement (iii) of Lemma 4 is satisfied for $j = 1$. The assumptions (ii) and (iii) of Lemma 4 for $j = 2$ can be easily checked. In fact we have

$$\Lambda_2(\xi) = (1-\nu) \int_{-\infty}^{\infty} \operatorname{sgn}(t-\xi) f_{\varepsilon,h}(t-\mu)dt + 2\nu \int_{-\infty}^{\infty} (t-\xi) f_{\varepsilon,h}(t-\mu)dt$$

$$= (1-\nu)[\int_{\xi}^{\infty} f_{\varepsilon,h}(t-\mu)dt - \int_{-\infty}^{\xi} f_{\varepsilon,h}(t-\mu)dt] + 2\nu(\mu-\xi) =$$

$$= (1-\nu)[1-2F_{\varepsilon,h}(\xi-\mu)] + 2\nu[\mu-\xi] \quad ,$$

which is a differentiable function in ξ, and hence $\Lambda_2(\xi)$ is differentiable at $\xi = \mu$. Further

$$(3.5) \qquad A_2 = \Lambda_2'(\xi)\Big|_{\xi=\mu} = -2[(1-\nu)f_{\varepsilon,h}(0)+\nu]$$

$$= -2[\frac{(1-\nu)}{\sqrt{2\pi}}[(1-\varepsilon)+\frac{\varepsilon}{h}] + \nu] < 0 \quad ,$$

which shows, that the requirement (ii) is satisfied. To show (iii) we have to evaluate the integral $I_2(\xi)$. We have

$$I_2(\xi) = \int_{-\infty}^{\infty} \psi_2^2(t-\xi)dF_{\varepsilon,h}(t-\mu)$$

$$= \int_{-\infty}^{\infty} [(1-\nu)^2[\mathrm{sgn}(t-\xi)]^2+4\nu^2(t-\xi)^2+4\nu(1-\nu)|t-\xi|]f_{\varepsilon,h}(t-\mu)dt$$

$$= (1-\nu)^2+4\nu^2 \int_{-\infty}^{\infty} (t-\xi)^2 f_{\varepsilon,h}(t-\mu)dt+4\nu(1-\nu)\int_{-\infty}^{\infty} |t-\xi|f_{\varepsilon,h}(t-\mu)dt$$

$$= (1-\nu)^2+4\nu^2[(1-\varepsilon)+\varepsilon h^2+(\mu-\xi)^2]+4\nu(1-\nu)[2(1-\varepsilon)\phi(\xi-\mu)+2h\varepsilon\phi\,(\frac{\xi-\mu}{h})]$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad .$$

From this it is easy to see that the integral $I_2(\xi)$ is finite and continuous at $\xi = \mu$ for $0 \le \varepsilon \le \frac{1}{2}$ and $1 < h \le h^* < \infty$ with h^* fixed. We have

$$(3.6)\qquad B_2 = \lim_{\xi \to \mu} \int_{-\infty}^{\infty} \psi_2^2(t-\xi)dF_{\varepsilon,h}(t-\mu) = \int_{-\infty}^{\infty} \psi_2^2(t-\mu)dF_{\varepsilon,h}(t-\mu)$$

$$= (1-\nu)^2+4\nu^2[(1-\varepsilon)+\varepsilon h^2]+8\nu(1-\nu)\frac{1}{\sqrt{2\pi}}[(1-\varepsilon)+\varepsilon h]\ .$$

IV.4 Mininax Solution

Now that the requirements of Lemma 4 are satisfied for each $T_\lambda^n \in E_1$, and $T_\nu^n \in E_2$, for fixed ε and h^*, $\sqrt{n}\,(T_\lambda^n-\mu)$ and $\sqrt{n}\,(T_\nu^n-\mu)$ are asymptotically normal with zero means and asymptotic variances

$$(4.1)\qquad V[\psi_1(t;\lambda),F_{\varepsilon,h}(t-\mu)] = \frac{\int_{-\infty}^{\infty}\psi_1^2(t-\mu)dF_{\varepsilon,h}(t-\mu)}{[\Lambda_1'(\mu)]^2} = \frac{B_1}{[A_1]^2}$$

$$= \frac{\sqrt{\pi}}{2}\ \frac{\Gamma(\lambda+\frac{1}{2})}{[\Gamma(1+\frac{\lambda}{2})]^2}\ \frac{[(1-\varepsilon)+\varepsilon h^{2\lambda}]}{[(1-\varepsilon)+\varepsilon h^{\lambda-1}]^2}\ ,$$

and

$$(4.2)\qquad V[\psi_2(t;\nu),F_{\varepsilon,h}(t-\mu)] = \frac{\int_{-\infty}^{\infty}\psi_2^2(t-\mu)dF_{\varepsilon,h}(t-\mu)}{[\Lambda_2'(\mu)]^2} = \frac{B_2}{[A_2]^2}$$

$$= \frac{\{(1-\nu)^2+4\nu^2[(1-\varepsilon)+\varepsilon h^2] + 8\nu(1-\nu)[\frac{1}{\sqrt{2\pi}}][(1-\varepsilon)+\varepsilon h]\}}{4\{(1-\nu)[\frac{1}{\sqrt{2\pi}}][(1-\varepsilon)+\frac{\varepsilon}{h}]+\nu\}^2}\ ,$$

where A_1, B_1, A_2, B_2 are defined by (3.3), (3.4), (3.5), (3.6), respectively.

THEOREM 1: There exists a unique most robust estimate T_λ^n in the class E_1, and a unique most robust estimate T_ν^n in the class E_2, corresponding to the choice of parameters $\lambda = \lambda^*(\varepsilon) \in [0,1]$ and $\nu = \nu^*(\varepsilon) \in [0,1]$, which minimize the suprema of the asymptotic variances $V[\psi_1(t,\lambda), F_{\varepsilon,h}(t-\mu)]$ and $V[\psi_2(t;\nu), F_{\varepsilon,h}(t-\mu)]$ respectively, the suprema being taken over the set

$$(4.3) \qquad F = \{F_{\varepsilon,h}(t-\mu) : F_{\varepsilon,h}(t-\mu) = (1-\varepsilon)\Phi(t-\mu)+\varepsilon\Phi(\frac{t-\mu}{h}) , \quad 1 < h \leq h^* < \infty\}$$

of all possible underlying distributions for the one-sample problem of estimating a single location parameter μ $(-\infty < \mu < \infty)$, where Φ is the standard normal distribution function and $0 \leq \varepsilon \leq \frac{1}{2}$ and h^* are fixed.

PROOF: It will be convenient to distinguish two cases, the case 1, $\varepsilon \in (0, \frac{1}{2}]$ and the case 2, $\varepsilon = 0$.

Case 1. Suppose $\varepsilon \in (0, \frac{1}{2}]$. It is then seen that both asymptotic variances (4.1) and (4.2) are increasing functions of the scale ratio h, since the numerator and the denominator of the expression for the asymptotic variances are non-decreasing and non-increasing functions of h respectively for $0 \leq \lambda, \nu \leq 1$ and $0 < \varepsilon \leq \frac{1}{2}$ fixed. Since the distributions $F_{\varepsilon,h}(t-\mu)$ range over the set F defined by (4.3), the superma of the asymptotic variances over the set F will be attained by the distribution function with the greatest possible scale ratio $h = h^* < \infty$, i.e.,

$$(4.4) \qquad \sup_{1<h\leq h^*<\infty} V[\psi_1(t;\lambda),F_{\varepsilon,h}(t-\mu)) = V[\psi_1(t;\lambda),F_{\varepsilon,h=h*}(t-\mu)]$$

for all $0 \leq \lambda \leq 1$, $0 < \varepsilon \leq \frac{1}{2}$, and

$$(4.5) \qquad \sup_{1<h\leq h^*<\infty} V[\psi_2(t;\nu),F_{\varepsilon,h}(t-\mu)] = V[\psi_2(t;\nu),F_{\varepsilon,h=h*}(t-\mu)]$$

for all $0 \leq \nu \leq 1$, $0 < \varepsilon \leq \frac{1}{2}$.

Since our measure of robustness is the inverse of the supremum of the asymptotic variance over the set F of all possible underlying distributions, the most robust estimates in the classes E_1 and E_2 will correspond to the choices of parameters $\lambda = \lambda^*(\varepsilon) \in [0,1]$ and $\nu = \nu^*(\varepsilon) \in [0,1]$, which will minimize the suprema of asymptotic variances.

Let us denote by $W_2(\nu)$, the supremum of the asymptotic variance (4.2) over the set F, i.e.,

$$W_2(\nu) = \frac{(1-\nu)^2+4\nu^2A+8\nu(1-\nu)B}{4[(1-\nu)C+\nu]^2} = \frac{\nu^2[1+4A-8B]+2\nu[4B-1]+1}{4[\nu(1-C)+C]^2}$$

where

$$A = [(1-\varepsilon)+\varepsilon h^{*2}], \qquad B = [(1-\varepsilon)+\varepsilon h^*]\frac{1}{\sqrt{2\pi}}, \qquad C = [(1-\varepsilon)+\frac{\varepsilon}{h^*}]\frac{1}{\sqrt{2\pi}}.$$

The function $W_2(\nu)$ is differentiable in ν on $[0,1]$. It can be shown that

$$\frac{d}{d\nu} W_2(\nu) = \frac{4\{(4BC-1) - \nu[(4BC-1) + (4B-4AC)]\}}{8\{[\nu(1-C)+C]\}^3},$$

where

$$(4BC-1) = (\frac{2}{\pi} - 1) + \frac{2}{\pi}\frac{\varepsilon(1-\varepsilon)}{h^*}(h^*-1)^2$$

$$(4B-4AC) = \frac{4\varepsilon(1-\varepsilon)}{\sqrt{2\pi}}\frac{(h^*-1)(1-h^{*2})}{h^*}$$

$$[(4BC-1)+(4B-4AC)] = \frac{2\varepsilon(1-\varepsilon)}{\pi}\left[\frac{(h^*-1)^2}{h^*} + \sqrt{2\pi}\,\frac{(h^*-1)(1-h^{*2})}{h^*}\right]+(\frac{2}{\pi}-1) \ .$$

It is easy to see that

$$[(4BC-1)+(4B-4AC)] < 0 \quad \text{and} \quad (4B-4AC) < 0$$

for all $\varepsilon \in (0, \frac{1}{2}]$ and all $1 < h^* < \infty$, and

$$8\{[\nu(1-C)+C]\}^3 > 0$$

for all $\nu \in [0,1]$, $\varepsilon \in (0, \frac{1}{2}]$ and all $1 < h^* < \infty$. Hence

$$(4BC-1) > 0 \Longrightarrow \frac{d}{d\nu} W_2(\nu) > 0 \quad \text{on } [0,1] \ ,$$

and

$$(4BC-1) \leqq 0 \quad \text{and} \quad \nu = \nu^*(\varepsilon) = \frac{(4BC-1)}{(4BC-1)+(4B-4AC)} \in [0,1]$$

$$\Longrightarrow \frac{d}{d\nu} W_2(\nu)\Big|_{\nu=\nu^*(\varepsilon)} = 0 \ .$$

Further

$$(4BC-1) + C(4B-4AC) < 0 \Longrightarrow \frac{d^2}{d\nu^2} W_2(\varepsilon)\Big|_{\nu=\nu^*(\varepsilon)} > 0 \ .$$

Thus we have that the function $W_2(\nu)$ has an absolute minimum on $[0,1]$ at $\nu = \nu^*(\varepsilon) = 0$, if $(4BC-1) \geqq 0$, which is equivalent to the condition $h^* \geq h_2(\varepsilon)$, and the function $W_2(\nu)$ has an absolute minimum at

$$\nu = \nu^*(\varepsilon) = \frac{(4BC-1)}{(4BC-1)+(4B-4AC)} = \frac{h^*(\pi-2) - 2\varepsilon(1-\varepsilon)(h^*-1)^2}{h^*(\pi-2)+2\varepsilon(1-\varepsilon)(h^*-1)^2[(h^*+1)\sqrt{2\pi}-1]}$$

if $(4BC-1) < 0$, which is equivalent to the condition $1 < h^* < h_2(\varepsilon)$, where $h_2(\varepsilon)$ (> 1) is given by

$$h_2(\varepsilon) = 1 + \frac{1}{2\varepsilon(1-\varepsilon)} [(\frac{\pi}{2}-1) + \sqrt{(\frac{\pi}{2}-1)[(\frac{\pi}{2}-1)+4\varepsilon(1-\varepsilon)]}]$$

for $\varepsilon \in (0, \frac{1}{2}]$.

Let us denote by $W_1(\lambda)$ the supermum of the asymptotic variance (4.1) over the set F, i.e.,

$$W_1(\lambda) = \frac{\sqrt{\pi}}{2} \frac{\Gamma(\lambda+\frac{1}{2})}{[\Gamma(1+\frac{\lambda}{2})]^2} \frac{[(1-\varepsilon)+\varepsilon h^{*2\lambda}]}{[(1-\varepsilon)+\varepsilon h^{*\lambda-1}]^2}$$

We shall show, that the function $W_1(\lambda)$ is a convex function on $[0,1]$ for all $\varepsilon \in (0, \frac{1}{2}]$, $1 < h^* < \infty$. Since logarithmic convexity implies convexity, it is enough to show that the function $\log W_1(\lambda)$ is a convex function on $[0,1]$. We have

$$\frac{d^2}{d\lambda^2}[\log W_1(\lambda)] =$$

$$= \frac{\Gamma''(\lambda+\frac{1}{2})\,\Gamma(\lambda+\frac{1}{2})-[\Gamma'(\lambda+\frac{1}{2})]^2}{[\Gamma(\lambda+\frac{1}{2})]^2} - \frac{1}{2}\,\frac{\Gamma''(\frac{\lambda}{2}+1)\Gamma(\frac{\lambda}{2}+1)-[\Gamma'(\frac{\lambda}{2}+1)]^2}{[\Gamma(\frac{\lambda}{2}+1)]^2}$$

$$+ \frac{2\varepsilon(1-\varepsilon)[\log(h^*)]^2}{[(1-\varepsilon)+\varepsilon h^{*2\lambda}]^2[(1-\varepsilon)+\varepsilon h^{*\lambda-1}]^2} \{(1-\varepsilon)^2[2h^{*2\lambda}-h^{*\lambda-1}]$$

$$+ 2\varepsilon(1-\varepsilon)h^{*3\lambda-1}+\varepsilon^2[2h^{*4\lambda-2}-h^{*5\lambda-1}]\}$$

from which by using the fact (see for example [15], p. 250) that

$$\frac{d^2}{dz^2} \log \Gamma(z) = \frac{\Gamma''(z).\Gamma(z)-[\Gamma'(z)]^2}{[\Gamma(z)]^2} = \sum_{n=o}^{\infty} \frac{1}{(n+z)^2}$$

we have

$$\frac{d^2}{d\lambda^2} [\log W_1(\lambda)] > 0 \quad \text{on} \quad [0,1] \quad ,$$

and hence $W_1(\lambda)$ is a convex function on $[0,1]$ for all $\lambda \in [0,1]$, $\varepsilon \in (0, \frac{1}{2}]$ and $1 < h^* < \infty$. Next

$$\frac{d}{d\lambda} [\log W_1(\lambda)] = \frac{\Gamma'(\lambda+\frac{1}{2})}{\Gamma(\lambda+\frac{1}{2})} - \frac{\Gamma'(\frac{\lambda}{2}+1)}{\Gamma(\frac{\lambda}{2}+1)}$$

$$+ \left\{\frac{2\varepsilon(1-\varepsilon)\log (h^*)[h^{*2\lambda}-h^{*\lambda-1}]}{[(1-\varepsilon)+\varepsilon h^{*2\lambda}][(1-\varepsilon)+\varepsilon h^{*\lambda-1}]}\right\}$$

$$= \{K(\lambda)\} + \{L_\varepsilon(\lambda)\} \text{ , say.}$$

Since the derivatives of the functions $K(\lambda)$ and $L_\varepsilon(\lambda)$ are strictly positive on $[0,1]$, both $K(\lambda)$ and $L_\varepsilon(\lambda)$ are strictly increasing functions on $[0,1]$. Further $K(\lambda) \leq 0$ on $[0,1]$ and $L_\varepsilon(\lambda) > 0$ on $[0,1]$. From this it is easy to see, that

$$\min_\lambda K(\lambda) + \min_\lambda L_\varepsilon(\lambda) > 0 \implies \frac{d}{d\lambda} [\log W_1(\lambda)] > 0 \quad \text{on} \quad [0,1] \quad ,$$

and

$$\min_\lambda K(\lambda) + \min_\lambda L_\varepsilon(\lambda) \leq 0 \implies \frac{d}{d\lambda} [\log W_1(\lambda)] = 0$$

for some $\lambda = \lambda^*(\varepsilon) \in [0,1]$.

Thus we have, that the function $W_1(\lambda)$ has an absolute minimum on $[0,1]$ at $\lambda = \lambda^*(\varepsilon) = 0$ if

$$\min_{\lambda} K(\lambda) + \min_{\lambda} L_{\varepsilon}(\lambda) \geqq 0 \quad ,$$

which is equivalent to the condition $h^* \geq h_1(\varepsilon)$, and the function $W_1(\lambda)$ has an absolute minimum on $[0,1]$ at $\lambda = \lambda^*(\varepsilon) \in (0,1]$ if

$$\min_{\lambda} K(\lambda) + \min_{\lambda} L_{\varepsilon}(\lambda) < 0 \quad ,$$

which is equivalent to the condition $1 < h^* < h_1(\varepsilon)$, where $h_1(\varepsilon)$ (> 1) is given as a solution of

$$\frac{2\varepsilon(1-\varepsilon)[1 - \frac{1}{h_1(\varepsilon)}] \log [h_1(\varepsilon)]}{[1-\varepsilon(1 - \frac{1}{h_1(\varepsilon)})]} = \log 4 \quad \text{for } \varepsilon \in (0,\tfrac{1}{2}] \ .$$

Case 2: Suppose $\varepsilon = 0$. Then the class of underlying distributions F reduces to a single member $F_{o,h}(t-\mu) = \Phi(t-\mu)$. The asymptotic variances (4.1) and (4.2) are then of the following forms:

$$V[\psi_1(t;\lambda),F_{o,h}(t-\mu)] = \frac{\sqrt{\pi}}{2} \frac{\Gamma(\lambda+\frac{1}{2})}{[\Gamma(\frac{\lambda}{2}+1)]^2}$$

and

$$V[\psi_2(t;\nu),F_{o,h}(t-\mu)] = \frac{(1-\nu)^2 + 4\nu^2 + 8\nu(1-\nu)[\frac{1}{\sqrt{2\pi}}]}{4\{(1-\nu)[\frac{1}{\sqrt{2\pi}}] + \nu\}^2} \ .$$

It is easy to see that the asymptotic variances are minimized in both instances for $\lambda = \nu = 1$. This proves the theorem.

Remark: From the above theorem we have

In the case when no contamination is present, i.e., when $\varepsilon = 0$, the most robust estimates for the location parameter μ in the classes E_1 and E_2 correspond to $T^n_{\lambda=1} = T^n_{\nu=1}$, which is the sample mean.

In the case when $\varepsilon \in (0, \frac{1}{2}]$, the most robust estimate for the location parameter μ in the class E_1 corresponds to $T^n_{\lambda=o}$, i.e., to the sample median if the asymptotically least favourable distribution in the class F has scale ratio $h^* \geq h_1(\varepsilon)$. This implies that the most robust estimate in E_1 cannot correspond to the sample median if the scale ratio

$$1 < h^* < \inf_{0<\varepsilon\leq\frac{1}{2}} h_1(\varepsilon) \doteq 6.6 \quad .$$

In the case when $\varepsilon \in (0, \frac{1}{2}]$, the most robust estimate for the location parameter μ in the class E_2 corresponds to $T^n_{\nu=o}$, i.e., to the sample median if the asymptotically least favourable distribution in the class F has scale ratio $h^* \geq h_2(\varepsilon)$. This implies that the most robust estimate in E_2 cannot correspond to the sample median if the scale ratio

$$1 < h^* < \inf_{0<\lambda\leq\frac{1}{2}} h_2(\varepsilon) \doteq 4.0 \quad .$$

In the case when $\varepsilon \in (0, \frac{1}{2}]$, and when the scale ratio h^* of the asymptotically least favourable distribution in F lies between $1 < h^* < h_1(\varepsilon)$, the most robust estimate for the location parameter μ corresponds to the choice of parameter $\lambda = \lambda^*(\varepsilon)$ in the interval $[0,1)$.

In the case when $\varepsilon \in (0, \frac{1}{2}]$, and when the scale ratio h^* of the asymptotically least favourable distribution in $\mathcal{F}$ lies between $1 < h^* < h_2(\varepsilon)$, the most robust estimate for the location parameter μ in E_2 corresponds to the choice of parameter $\nu = \nu^*(\varepsilon)$ in the interval $[0,1)$.

Thus if the scale ratio h^* of the asymptotically least favourable distribution in $\mathcal{F}$,

$$h^* \in (1, \inf_{o<\lambda\leq\frac{1}{2}} h_2(\varepsilon) \doteq 4.0) \quad ,$$

then the most robust estimate in any of the classes E_1 or E_2 cannot correspond to the sample median.

Suppose now $h^* = 3$. Then the class $\mathcal{F}$ of underlying distributions for the one sample problem of estimating a location parameter μ reduces to

$$(4.11) \qquad \overline{\mathcal{F}} = F_{\varepsilon,h}(t-\mu) : F_{\varepsilon,h}(t-\mu) = (1-\varepsilon)\Phi(t-\mu)+\varepsilon\Phi(\frac{t-\mu}{h}) \;, \quad 1 < h \leqq h^* = 3\} \quad .$$

The suprema of asymptotic variances of the estimates $T^n_\lambda \in E_1$ and $T^n_\nu \in E_2$ are attained in both instances for an asymptotically least favourable distribution function in $\overline{\mathcal{F}}$, corresponding to the scale ratio $h = h^* = 3$, i.e., when

$$F_{\varepsilon,h}(t-\mu) = F_{\varepsilon,h*=3}(t-\mu) = (1-\varepsilon)\Phi(t-\mu)+\varepsilon\Phi(\frac{t-\mu}{3})$$

is the Tukey's contaminated normal distribution.

The most robust estimate in E_2 corresponds to the choice of parameter $\nu = \nu^*(\varepsilon)$, given by

$$(4.12) \qquad \nu^*(\varepsilon) = \frac{3(\pi-2) - 8\varepsilon(1-\varepsilon)}{3(\pi-2) + 8\varepsilon(1-\varepsilon)[4\sqrt{2\pi} - 1]} \quad ,$$

for $\varepsilon \in [0,1]$, and the most robust estimate in E_1 corresponds to the choice of parameter $\lambda = \lambda^*(\varepsilon)$ given as a solution of

$$(4.13) \qquad \frac{d}{d\lambda} \{V[\psi_1(t;\lambda), F_{\varepsilon,h^*=3}(t-\mu)]\} = 0 \quad ,$$

for $\varepsilon \in [0,1]$, from which no easy explicit expression for $\lambda^*(\varepsilon)$ can be obtained.

From expression (4.12) it can be seen, that for the amount of contamination $\varepsilon = 0$, the best choice of ν is $\nu = \nu^*(\varepsilon=o) = 1$, i.e., the most robust estimate for the location parameter μ corresponds to the sample mean $T^n_{\nu=1}$. For the maximal amount of contamination $\varepsilon = \frac{1}{2}$, the best choice of ν is $\nu = \nu^*(\varepsilon = \frac{1}{2}) \doteq 0.0663$, i.e., the most robust estimate for the location parameter μ corresponds to the estimate $T^n_{\nu\doteq 0.0663}$.

The following two tables show the suprema of the asymptotic variances of the estimates T^n_ν and T^n_λ, the suprema being taken over the set $\overline{F}$ defined by (4.11) - for selected values of λ,ν and amounts of contamination ε, i.e., the asymptotic variances correspond to the asymptotically least favourable distribution in the set $\overline{F}$, which is the Tukey's contaminated normal distribution with scale ratio $h^* = 3$. The values with asterisks correspond to the minima of asymptotic variances for different amounts of contamination $\varepsilon \in [0, \frac{1}{2}]$, attained by the estimates $T^n_{\lambda^*(\varepsilon)}$ and $T^n_{\nu^*(\varepsilon)}$.

TABLE 1

The Supremum Over the Set $\mathcal{F}$ of the Asymptotic Variance of the Estimate T_ν^n Defined as a Solution of

$$(1-\nu)\sum_{i=1}^{n}|T_\nu^n - X_i| + \nu\sum_{i=1}^{n}(T_\nu^n - X_i)^2 = \text{minimum},$$

$$0 \leqq \nu \leqq 1, \quad 0 \leqq \varepsilon \leqq \frac{1}{2}$$

ν \ ε	0.00	0.01	0.05	0.10	0.20	0.30	0.40	0.50
0.00	1.571	1.592	1.681	1.803	2.091	2.454	2.921	3.534
0.01	1.543	1.564	1.654	1.777	2.065	2.428	2.892	3.499
0.05	1.445	1.468	1.562	1.691	1.991	2.363	2.830*	3.427*
0.10	1.349	1.374	1.478	1.620	1.946*	2.342*	2.832	3.439
0.20	1.216	1.247	1.378	1.554	1.949	2.413	2.961	3.612
0.30	1.133	1.172	1.332	1.545*	2.012	2.543	3.148	3.840
0.40	1.080	1.126	1.316	1.564	2.098	2.688	3.341	4.065
0.50	1.046	1.100	1.316*	1.597	2.190	2.831	3.523	4.272
0.60	1.025	1.085	1.326	1.636	2.282	2.966	3.689	4.455
0.70	1.012	1.078	1.342	1.678	2.370	3.090	3.839	4.617
0.80	1.005	1.075*	1.360	1.720	2.453	3.203	3.972	4.760
0.90	1.001	1.077	1.380	1.761	2.529	3.306	4.092	4.887
1.00	1.000*	1.080	1.400	1.800	2.600	3.400	4.200	5.000

TABLE 2

The Supremum Over the Set $\overline{F}$ of the Asymptotic Variance of the Estimate T_λ^n Defined as a Solution of

$$\sum_{i=1}^{n} |T^n - X_i|^\lambda \operatorname{sgn}(T_\lambda^n - X_i) = 0 ,$$

$$0 \leqq \lambda \leqq 1, \quad 0 \leqq \varepsilon \leqq \frac{1}{2}$$

λ \ ε	0.00	0.01	0.05	0.10	0.20	0.30	0.40	0.50
0.00	1.571	1.592	1.681	1.803	2.091	2.454	2.921	3.534
0.01	1.549	1.571	1.659	1.781	2.069	2.430	2.895	3.506
0.05	1.473	1.494	1.582	1.703	1.989	2.348	2.808	3.409
0.10	1.393	1.414	1.503	1.624	1.911	2.270	2.728	3.323
0.20	1.271	1.293	1.386	1.513	1.810	2.179	2.644	3.239*
0.30	1.185	1.209	1.310	1.447	1.765	2.155*	2.639*	3.246
0.40	1.123	1.150	1.263	1.415	1.764*	2.185	2.697	3.326
0.50	1.079	1.110	1.238	1.411*	1.802	2.264	2.812	3.468
0.60	1.047	1.083	1.233*	1.433	1.877	2.389	2.980	3.668
0.70	1.025	1.068	1.247	1.481	1.991	2.562	3.202	3.921
0.80	1.010	1.063*	1.278	1.556	2.146	2.785	3.477	4.228
0.90	1.002	1.067	1.328	1.661	2.347	3.063	3.809	4.587
1.00	1.000*	1.080	1.400	1.800	2.600	3.400	4.200	5.000

From Tables 1 and 2, $T^n_{\lambda*(\varepsilon)}$ seems to be preferable to $T^n_{\nu*(\varepsilon)}$ for all $\varepsilon \in [0, \frac{1}{2}]$.

From a practial point of view it is unpleasant that the most robust estimate depends on the amount of contamination ε, but it is an unavoidable difficulty. If we assume, that contamination can occur in the whole range $\varepsilon \in [0, \frac{1}{2}]$, then from Tables 1 and 2 the estimates $T^n_{\nu=\frac{1}{4}}$ and $T^n_{\lambda=\frac{1}{2}}$ seem to be reasonable choices. If we assume, that contamination can occur in the smaller range $\varepsilon \in [0, \frac{1}{10}]$, then from Tables 1 and 2 the estimates $T^n_{\nu=\frac{1}{2}}$ and $T^n_{\lambda=\frac{2}{3}}$ seem to be reasonable choices.

The averages of functions $\nu*(\varepsilon)$ and $\lambda*(\varepsilon)$ over the intervals $[0, \frac{1}{2}]$ and $[0, \frac{1}{10}]$ yield the following results, where the values for $\lambda*(\varepsilon)$ were obtained by a graphical method:

$$\nu_o = 2\int_0^{\frac{1}{2}} \nu*(\varepsilon)d\varepsilon \doteq \underline{0.1922}; \qquad \nu_1 = \int_0^{\frac{1}{10}} \nu*(\varepsilon)d\varepsilon \doteq \underline{0.4999};$$

$$\lambda_o = 2\int_0^{\frac{1}{2}} \lambda*(\varepsilon)d\varepsilon \doteq \underline{0.45}; \qquad \lambda_1 = \int_0^{\frac{1}{10}} \lambda*(\varepsilon)d\varepsilon \doteq \underline{0.65};$$

Hence the estimates $T^n_{\nu_o}$, $T^n_{\lambda_o}$ and $T^n_{\nu_1}$, $T^n_{\lambda_1}$ could be also reasonable candidates if the occurrence of contamination ε is in the ranges $[0, \frac{1}{2}]$ and $[0, \frac{1}{10}]$ respectively.

Now we shall turn to computational aspects of the estimates T^n_ν and T^n_λ. First we consider the estimate T^n_ν.

<u>LEMMA 1</u>: Let $x = (x_1,\ldots,x_n)$ be an arbitrary sample point and let T^n_ν be the estimate defined for $0 \leq \nu \leq 1$ by:

$$Q_\nu(c) = (1-\nu) \sum_{i=1}^{n} |x_i - c| + \nu \sum_{i=1}^{n} (x_i - c)^2 = \min \iff c = T_\nu^n \quad .$$

Then for each $0 \leq \nu \leq 1$ we have

$$T_\nu^n \in [T_{\nu=o}^n, T_{\nu=1}^n] = [\tilde{x}, \bar{x}] \quad ,$$

where $\tilde{x}$ and $\bar{x}$ are the sample median and the sample mean respectively.

<u>PROOF</u>: It is seen that $Q_\nu(c)$ is for each fixed $\nu \in [0,1]$ a convex function in c.

For $\nu = 0$ and $\nu = 1$ we have

$$(4.14) \qquad Q_o(c) = \sum_{i=1}^{n} |x_i - c| = \min \iff c = T_{\nu=o}^n = \tilde{x}$$

and

$$(4.15) \qquad Q_1(c) = \sum_{i=1}^{n} |x_i - c|^2 = \min \iff c = T_{\nu=1}^n = \bar{x}$$

respectively. From the definition of the estimate T_ν^n we have for each $0 \leq \nu \leq 1$

$$Q_\nu(T_{\nu=o}^n) = Q\ (\tilde{x}) \geq Q_\nu(T_\nu^n) = \min_c Q_\nu(c)$$

and

$$Q_\nu(T_{\nu=1}^n) = Q_\nu(\bar{x}) \geq Q_\nu(T_\nu^n) = \min_c Q_\nu(c) \quad ,$$

hence if $a = \min(\bar{x}, \tilde{x})$, $b = \max(\bar{x}, \tilde{x})$, we have for each $0 \leq \nu \leq 1$

$$Q_\nu(a) \geq Q_\nu(T^n_\nu) \qquad \text{and} \qquad Q_\nu(T^n_\nu) \leq Q_\nu(b) \quad .$$

Thus in order to show that $T^n_\nu \in [T^n_{\nu=o}, T^n_{\nu=1}]$ for each $0 \leq \nu \leq 1$, it is enough to show that

$$Q_\nu(a-\Delta) \geq Q_\nu(a) \qquad \text{and} \qquad Q_\nu(b) \leq Q_\nu(b+\Delta)$$

for each $\Delta > 0$. We can write $Q_\nu(c)$ in the form

$$Q_\nu(c) = (1-\nu) \sum_{i=1}^{n} |x_i - c| + \nu \sum_{i=1}^{n} (x_i - c)^2$$

$$= (1-\nu)Q_o(c) + \nu Q_1(c) \quad .$$

It is convenient to distinguish two situations. First let $a = \bar{x}$ and $b = \tilde{x}$. Then from the convexity of $Q_\nu(c)$, $Q_o(c)$ and $Q_1(c)$, and from the fact that $\bar{x}-\Delta < \bar{x} \leqq \tilde{x} < \tilde{x}+\Delta$ we have

$$Q_\nu(a) = Q_\nu(\bar{x}) = (1-\nu)Q_o(\bar{x})+\nu Q_1(\bar{x}) \leqq (1-\nu)Q_o(\bar{x})+\nu Q_1(\bar{x}-\Delta)$$

$$\leqq (1-\nu)Q_o(\bar{x}-\Delta)+\nu Q_1(\bar{x}-\Delta) = Q_\nu(\bar{x}-\Delta) = Q_\nu(a-\Delta)$$

and

$$Q_\nu(b) = Q_\nu(\tilde{x}) = (1-\nu)Q_o(\tilde{x})+\nu Q_1(\tilde{x}) \leqq (1-\nu)Q_o(\tilde{x}+\Delta)+\nu Q_1(\tilde{x})$$

$$\leqq (1-\nu)Q_o(\tilde{x}+\Delta)+\nu Q_1(\tilde{x}+\Delta) = Q_\nu(\tilde{x}+\Delta) = Q_\nu(b+\Delta) \quad .$$

For the case $a = \tilde{x}$ and $b = \bar{x}$ by the completely analogous argument we can show

$$Q_\nu(a) = Q_\nu(\tilde{x}) = (1-\nu)Q_o(\tilde{x})+\nu Q_1(\tilde{x}) \leqq (1-\nu)Q_o(\tilde{x}-\Delta)+\nu Q_1(\tilde{x})$$

$$\leqq (1-\nu)Q_o(\tilde{x}-\Delta)+\nu Q_1(\tilde{x}-\Delta) = Q_\nu(\tilde{x}-\Delta) = Q_\nu(a-\Delta)$$

and

$$Q_\nu(b) \leqq Q_\nu(b+\Delta) \quad .$$

Thus we have $Q_\nu(a-\Delta) \geq Q_\nu(a)$ and $Q_\nu(b) \leqq Q_\nu(b+\Delta)$, which terminates the proof.

<u>LEMMA 2</u>: Let $x = (x_1,\ldots,x_n)$ be an arbitrary sample point, and $y_1 < y_2 < \ldots < y_n$ be the corresponding ordered sample point. Consider intervals I_i for $i = 0,1,\ldots,n$ where $I_i = (y_i,y_{i+1}]$ for $i = 1,2,\ldots,(n-1)$, and $I_o = (-\infty, y_1]$, $I_n = (y_n,\infty)$.

If the following two conditions

$$y_i < c^*(i) = \bar{x} + \frac{[n-2i](1-\nu)}{2n\nu} \tag{4.16}$$

and

$$y_{i+1} \geq c^*(i+1) = \bar{x} + \frac{[n-2(i+1)](1-\nu)}{2n.\nu} \tag{4.17}$$

are satisfied, then

$$\begin{aligned} T^n_\nu &= c^*(i) = \bar{x} + \frac{[n-2i][1-\nu]}{2n\nu} && \text{for } c^*(i) \in (y_i,y_{i+1}] \ , \\ &= y_{i+1} && \text{for } c^*(i) \notin (y_i,y_{i+1}] \ , \end{aligned}$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad ,$$

and $\nu \in (0,1)$ (excluding the case of $\nu = 0$ and $\nu = 1$, which correspond to sample median and sample mean respectively).

PROOF: From the definition of T_ν^n we have for $0 \leqq \nu \leqq 1$:

$$Q_\nu(c) = (1-\nu)\sum_{j=1}^{n} |c-x_j| + \nu \sum_{j=1}^{n} (c-x_j)^2 = \min \Longleftrightarrow c = T_\nu^n$$

or equivalently for ordered x_i's:

$$Q_\nu(c) = (1-\nu)\sum_{j=1}^{n} |c-y_j| + \nu \sum_{j=1}^{n} (c-y_j)^2 = \min \Longleftrightarrow c = T_\nu^n \quad .$$

Consider the function $Q_\nu(c)$ on an arbitrary interval I_i for $i = 0,1,\ldots,n$, and for $0 < \nu < 1$.

For $c \in I_i$, $i = 0,1,\ldots,n$ and $0 < \nu < 1$ we have

$$(4.18) \qquad Q_\nu(c) = (1-\nu)[\{\textstyle\sum_{j=1}^{n} y_j - 2\sum_{j=1}^{i} y_j\} - c[n-2i]]$$

$$+ \nu\{nc^2 - 2c\sum_{j=1}^{n} y_j + \sum_{j=1}^{n} y_j^2\}$$

$$= [n\nu]c^2 - \{(n-2i)(1-\nu) + 2\nu\sum_{j=1}^{n} y_j\} c$$

$$+ \{\nu\sum_{j=1}^{n} y_j^2 + (1-\nu)[\sum_{j=1}^{n} y_j - 2\sum_{j=1}^{i} y_j]\} \quad .$$

The derivative of the function $Q_\nu(c)$ in the interior of the interval I_i and the right derivative at the leftmost point and the left derivative at the rightmost point are given by

$$(4.19) \qquad \frac{d}{dc} Q_\nu(c) = 2n\nu c - \{(n-2i)(1-\nu) + 2\nu\sum_{j=1}^{n} y_j\} \quad .$$

If the convex function $Q_\nu(c)$ has minimum in the interior of the interval I_i at the point $c^*(i)$ for $i = 1,2,\ldots,(n-1)$, then

$$\frac{d}{dc} Q_\nu(c) \Big|_{c=c^*(i)} = 0 \quad , \tag{4.20}$$

i.e.,

$$c^*(i) = \frac{(n-2i)(1-\nu)+2\nu\sum_{j=1}^{n} y_j}{2n\nu} = \bar{x} + \frac{(n-2i)(1-\nu)}{2n\nu} \quad .$$

The necessary and sufficient condition for the function $Q_\nu(c)$ to have minimum in the interval $I_i = (y_i, y_{i+1}]$, $i = 1,2,\ldots,(n-1)$, is that the right derivative of the function $Q_\nu(c)$ on the interval I_i at the leftmost point is negative and the right derivative of the function $Q_\nu(c)$ on the interval I_{i+1} at the leftmost point is non-negative.

The following two inequalities obtained from (4.20) express the above fact.

$$y_i < c^*(i) = \bar{x} + \frac{[(n-2i)(1-\nu)]}{2\nu n} \quad ,$$

$$y_{i+1} \geq c^*(i+1) = \bar{x} + \frac{[n-2(i+1)](1-\nu)}{2\nu n} \quad .$$

From the differentiability of the function $Q_\nu(c)$ on the interior of intervals I_i, $i = 0,1,\ldots,n$ it then follows, that if for an interval I_i, $i = 1,2,\ldots,(n-1)$ (4.16) and (4.17) are satisfied, then

$$\begin{aligned} T_\nu^n &= c^*(i) = \bar{x} + \frac{(n-2i)(1-\nu)}{2n\nu} \quad , && \text{if} \quad c^*(i) \in (y_i, y_{i+1}] \\ &= y_{i+1} && \text{if} \quad c^*(i) \notin (y_i, y_{i+1}] \quad . \end{aligned}$$

Example 1: Let the ordered sample values for $n = 7$ be

$$(y_1, y_2, y_3, y_4, y_5, y_6, y_7) = (2,3,4,6,7,8,33) .$$

Then $\bar{x} = 9$ and $\tilde{x} = 6$.

From Lemma 1 we know that for $\nu \in (0,1)$ $T_\nu^n \in [\tilde{x}, \bar{x}]$. Let us find the value of the estimate $T_{\nu=0.1}^{n=7}$. For interval $(y_5, y_6] = (7,8]$ we have

$$c^*(5) = \bar{x} + \frac{(n-2i)(1-\nu)}{2\nu n} = 7\ \frac{1}{14} \quad ; \qquad c^*(6) = 5\ \frac{11}{14} .$$

The conditions (4.16) and (4.17) are satisfied for the interval $(y_5, y_6]$, since $7 < 7\frac{1}{14}$ and $8 \geq 5\frac{11}{14}$ hold. Since also $c^*(5) = 7\frac{1}{14} \in (7,8]$, we have $T_{\nu=0.1}^{n=7} = 7\frac{1}{14}$.

Example 2: For determination of $T_{\nu=0.05}^{n=7}$ for the same sample, let us consider interval $(y_4, y_5] = (6,7]$. We have

$$c^*(4) \doteq 7\ \frac{9}{14} \ ; \qquad c^*(5) = 4\ \frac{13}{14} \ .$$

The conditions (4.16) and (4.17) are satisfied for the interval $(6,7]$, since $6 < 7\frac{9}{14}$ and $7 \geq 4\frac{13}{14}$. Since $c^*(4) = 7\frac{9}{14} \notin (6,7]$, we have $T_{\nu=0.05}^{n=7} = 7 = y_5$.

Thus we showed, that with the help of Lemmas 1 and 2 the evaluation of the estimate T_ν^n is quite simple. Besides if the value of ν is fixed, then by evaluation of $c^*(i)$ for $i = 1,2,\ldots,n$, we can easily obtain the estimate T_ν^n .

Let us turn now to the estimate T_λ^n. If $x = (x_1, \ldots, x_n)$ is any sample point and $y_1 < y_2 < \ldots < y_n$ is the ordered sample point, the estimate was defined by

$$(4.21) \qquad Q_\lambda(c) = \sum_{j=1}^{n} |x_j - c|^\lambda \operatorname{sgn}(x_j - c) = 0 \iff c = T_\lambda^n \quad \text{for} \quad 0 \le \lambda \le 1 \ ,$$

or equivalently

$$(4.22) \qquad Q_\lambda(c) = \sum_{j=1}^{n} |y_j - c|^\lambda \operatorname{sgn}(y_j - c) = 0 \iff c = T_\lambda^n \quad \text{for} \quad 0 \le \lambda \le 1 \ .$$

To solve the equation (4.22) for some $0 < \lambda < 1$ there is no simple method and apparently an iteration must be used.

The obvious bounds for the solution (4.22) for any $0 \le \lambda \le 1$, are the maximum and the minimum value of the sample.

That the analogy of Lemma 1 does not hold for the estimate T_λ^n can be seen, by considering a sample point for which the sample median $\tilde{x}$ and sample mean $\bar{x}$ are equal. If the extimate T_λ^n would lie always between the sample mean and sample median it should then be equal to the value $\bar{x} = \tilde{x} = T_\lambda^n$ for any $0 < \lambda < 1$. But this is obviously not the case.

From the Tables 1 and 2 it can be seen that the estimate T_λ^n has generally lower asymptotic variances than the estimate T_ν^n. On the other hand the estimate T_ν^n is easier to compute and hence for a practical purpose the estimate T_ν^n would be more useful.

IV.5 Concluding Remarks

For a simple model of indeterminacy, in which for the special case $h^* = 3$, Tukey's prototype distribution function turned out to be the asymptotically least favourable distribution, we produced

two classes of estimates $T_\lambda^n \in E_1$ and $T_\nu^n \in E_2$ as intermediaries between sample mean and sample median.

We showed, using Huber's results that T_λ^n and T_ν^n are asymptotically normal for $0 \leq \lambda,\ \nu \leq 1$. On the basis of accepted measure of robustness of an estimate - the inverse of the supremum of the asymptotic variance of an estimate over the set of underlying distributions - we determined the most robust estimates of the unknown location parameter μ in the classes E_1 and E_2, which correspond to the choices of $\lambda = \lambda^*(\varepsilon)$ and $\nu = \nu^*(\varepsilon)$. On the whole $T^n_{\lambda^*(\varepsilon)}$ seems to be preferable to $T^n_{\nu^*(\varepsilon)}$ for all $\varepsilon \in [0, \frac{1}{2}]$. On the other hand $T^n_{\lambda^*(\varepsilon)}$ is harder to compute than $T^n_{\nu^*(\varepsilon)}$. Also $T^n_{\lambda^*(\varepsilon)}$ is both translation and scale invariant whereas $T^n_{\nu^*(\varepsilon)}$ is only translation invariant. $T^n_{\lambda^*(\varepsilon)}$ does not always lie between mean and median whereas $T^n_{\nu^*(\varepsilon)}$ does always lie between mean and median. We recommend the use of the above estimates whenever the contaminating distribution also is normal and an upper limit can be set to the ratio of the variance of the contaminating distribution to that of the contaminated distribution.

BIBLIOGRAPHY

[1] Bickel, P.J. (1965). On some robust estimates of location. Ann. Math. Statist. 36 847-858.

[2] Chernoff, H. and Savage, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric tests. Ann. Math. Statist. 29 972-994.

[3] Cramér, H. (1946). Mathematical Methods of Statistics. Princeton Univ. Press.

[4] Hodges, J.L. Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. Proc. Fifth Berkeley Symp. Math. Stat. Prob. I. 163-186.

[5] Hodges, J.L. Jr. and Lehmann E.L. (1963). Estimates of location based on rank tests. Ann. Math. Statist. 34 598-611.

[6] Huber, P.J. (1964). Robust estimation of location parameter. Ann. Math. Statist. 35 73-101.

[7] Huber, P.J. (1968). Robust estimation. Mathematical Centre Tracts 27 3-25. Mathematisch Centrum, Amsterdam.

[8] Jaeckel, L.A. (1971). Robust estimates of location: Symmetry and asymmetric contamination. Ann. Math. Statist. 42 1020-1034.

[9] LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. Univ. California Publ. Statist. 1 277-330.

[10] LeCam, L. (1958) Les propriétés asymptotiques des solutions de Bayes. Publications de l' Institut de Statistique de l' Université de Paris. 7 17-35.

[11] Loève, M. (1960). Probability Theory (2nd. ed.) Van Nostrand, New York.

[12] Puri, M.L. (1964). Asymptotic efficiency of a class of c- sample tests. Ann. Math. Statist. 35 102-121.

[13] Tukey, J.W. (1949). The simplest signed-rank tests. Mem. Report 17, Statistical Research Group, Princeton Univ.

[14] Tukey, J.W. (1960). A survey of sampling from contaminated distributions. Contributions to Probability and Statistics, Essays in honor of Harold Hotelling 448-485.

[15] Whittaker, E.T. and Watson G.N. (1946). A Course of Modern Analysis (4th ed.) Cambridge University Press.